NORTHWESTERN UNIVERSITY

A Key to Understanding Social Communication Deficits in Autism Spectrum Disorders: Neural Processing of Sound and Speech Intonation

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Neuroscience

By

Nicole Monique Russo

EVANSTON, ILLINOIS

June 2008

ABSTRACT

A Key to Understanding Social Communication Deficits in Autism Spectrum Disorders: Neural Processing of Sound and Speech Intonation

Nicole Monique Russo

Autism spectrum disorders (ASD) is a highly heterogeneous and pervasive developmental disorder that affects nearly 1 in 150 children. A primary indicator of ASD is behavioral language impairment with respect to social communication, but the neurophysiology behind this impairment is not well understood. Both the perception and production of prosody is impaired in children with ASD. A major acoustic cue of prosody is the variation in vocal pitch (fundamental frequency) contour which helps in the perception of emotion and intention when someone is speaking. This study investigated children with and without ASD in order to describe the relationship between the language impairment in ASD and speech-evoked auditory potentials (recorded from the brainstem and cortex) and auditory-vocal regulation of voice fundamental frequency. Brainstem responses were evaluated in response to a consonant-vowel syllable /da/ and fully voiced speech syllables /ya/ with variable pitch contours. Cortical data were collected in response to the speech syllable /da/. Finally, control of voice fundamental frequency was examined using an auditory feedback/pitch-shift reflex paradigm. This study is the first demonstration of abnormal speech-evoked auditory brainstem responses in children with ASD.

Further, it isolated a subset of children with ASD who exhibited brainstem deficits specific to pitch tracking, which may be related to behavioral problems with prosody. Cortical responses revealed that abnormal encoding of speech in quiet in children with ASD is comparable to the encoding of speech in background noise in TD children. Finally, this study demonstrated two mechanisms of abnormal audio-vocal system regulation of voice fundamental frequency in children with ASD. Relationships between physiology and core and receptive language abilities were also established. Taken together, these data provide objective measures of neural processing deficits in ASD which may be related to the severity of the language impairment and may inform future objective sub-classification of children on the autism spectrum. Further, due to the malleability of the brainstem and cortex, it is conceivable that these methods could be utilized in the identification of children who would be ideal candidates for auditory or music-related remediation programs or used to monitor the progress of children enrolled in such therapies.

ACKNOWLEDGEMENTS

These studies were supported by grants from National Institute of Health R01 DC01510 (Kraus) and R01 DC006243-02 (Larson). The author wishes to thank Nina Kraus, Ph.D. for encouragement to pursue the study of autism spectrum disorders even though this population was new for the laboratory and for the inspiration to always take the findings a step further. Thanks are also in order for committee members: Steve Zecker, Ph.D., for his wisdom about science and life and his statistical assistance; Barbara Trommer, M.D., for opening up her pediatric neurology clinic to a graduate student with little experience in autism spectrum disorders and for her devotion and assistance in helping the author to write and think more clearly; Ann Bradlow, Ph.D., for her forward thinking, calm nature, and linguistics expertise; Charles Larson, Ph.D., for the invitation into his laboratory and his persistence and efforts with respect to our collaborative project. In sum, I would like to thank all committee members for their commitment to this effort and their steadfast support. This work would not have been completed today had it not also been for the technical and emotional support and collaborative contributions of fellow members of both the Kraus and Larson laboratories: Trent Nicol, Erika Skoe, Hanjun Liu, Ph.D., Karen Banai, Ph.D., Erin Hayes, Ph.D., Gabriella Musacchia, Ph.D., Daniel Abrams, Krista Johnson, Judy Song, Jane Hornickel, Brad Wible, Ph.D., and Julia Chen. I would also like to acknowledge Robert Harper-Mangels, Ph.D., and past Assistant Director of NUIN, who has offered guidance and counseling well beyond the requirements of his position.

Finally, I offer special thanks to my enormous support system of family and friends who have been there for me in so many ways for the past seven years. It would be impossible to name

everyone here. However, I would like to especially acknowledge my mother, father, siblings Cherise, Chris, and Amy, my always supportive and understanding boyfriend Bruce, and my three year old niece Grace, whose voice on the other side of the telephone brightened my most challenging days. I hope that with the successes I have been afforded and for knowing that we have made a small difference in the lives of those affected by ASD, each of you is repaid (at least in part) for all of your love and support.

PREFACE

The following five studies were designed to assess the relationship between the language impairment in ASD and the neural processing of speech sounds. Chapter II establishes the basic methodology leading to the inception of the brainstem studies. Each study reported in Chapters III-VI compares responses in children with ASD to TD children, explores the relationship between language abilities and physiology, and attempts to explain the neurobiology behind deficits identified in children with ASD. Utilizing the methodology developed in Chapter II, Chapter III reports the fidelity of the brainstem response to a speech syllable /da/ in quiet and background noise. Developed as an outgrowth of Chapters II and III, Chapter IV focuses on brainstem transcription specifically with respect to pitch contour, a leading cue of prosody in speech. Chapter V expands upon the neural auditory processing deficits and characterizes the audio-vocal regulation of fundamental frequency (pitch). Finally, Chapter VI explores the cortical processing of /da/ in quiet and background noise. A summary of the findings across studies is provided in the Discussion (Chapter VII).

DEDICATION

I dedicate this dissertation to the children with autism spectrum disorders and their families who devoted many hours to the success of this study, sometimes coming from neighboring states and spending weekends in Chicago in order to participate. May this work have a lasting effect on the future study of autism spectrum disorders.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I: INTRODUCTION

The neurobiology and etiology of language and communication deficits in autism spectrum disorders (ASD) are poorly understood. ASD is a pervasive developmental disorder that affects nearly 1 in 150 individuals (Center for Disease Control Prevention, 2007), yet there are still no objective measures for diagnosis or prognosis. Instead, ASD is behaviorally characterized by 1) language impairment with respect to social communication; 2) repetitive or stereotyped behaviors or interests; and 3) impairment in social interaction. In each of these domains, the unifying feature is the aberrant social aspect of the behavior. Each core symptom may be affected to a varying degree of severity, resulting in a densely heterogeneous group of children who may or may not be presenting with disorders of similar origin (Filipek et al., 2000; Freitag, 2007; McGovern & Sigman, 2005; Salmond, Vargha-Khadem, Gadian, de Haan, & Baldeweg, 2007). Thus, in order to better understand the disorder, it is useful to investigate each core symptom individually.

The highly variable language impairment in ASD is the primary focus of this work. Some children are completely non-verbal. However, if and when language does develop, individuals with ASD may experience impairments in all domains of communication (e.g., expression, and perception of language), with the most severe impairments in the realm of social communication, also known as pragmatic language (Rapin & Dunn, 2003; Rutter, 1974). Prosody and the use of acoustic cues to convey emotion and intention are severely impaired in individuals with ASD (Hubbard & Trauner, 2007). Some children with ASD may experience a significant delay in language development or they may regress (and lose language) after having acquired language

(De Giacomo & Fombonne, 1998). Consistent with delayed language abilities, abnormal or absent babbling and overall deficient oral-motor coordination has been reported in infants with ASD (Dawson, Osterling, Meltzoff, & Kuhl, 2000; Gernsbacher, 2004; Rutter, 1974).

Motivation and Hypothesis

Successful social communication involves being able to perceive and produce sounds meaningfully. An unfortunate, yet common, theme to the study of ASD is the lack of objective tools for identification and classification of biological markers for the language impairment in this heterogeneous group of individuals. Thus, the goal of this work is to develop an understanding of the neurobiology of auditory and audio-vocal neural processing in ASD and the relationship to language and prosody-specific deficits, as well as to inform intervention strategies for children with ASD. In order to accomplish this goal, five experiments were planned. This comprehensive study investigates 1) the neural transcription of speech in quiet and background noise at the brainstem level in normal systems; 2) the neural transcription of speech at the brainstem level; 3) the brainstem transcription of variable pitch contour; 4) the audio-vocal system involvement with regulation of voice fundamental frequency (F_0); 5) the cortical encoding of speech in quiet and background noise. Experiments 2-5 were evaluated in both children with ASD and typically-developing (TD) control children. Finally, the relationships between these physiological measures with cognitive and behavioral measures were investigated. The central hypothesis for this study is that difficulties perceiving and producing speech and prosody (pitch) are due to disordered representations of speech in the brainstem and cortex. These independent, yet related, studies explore in greater detail the deficits in speech-related

central auditory processing and audio-vocal interactions in ASD. However, it is necessary to begin by briefly introducing each component in both normal and language-disordered systems.

Prosody: What is it?

Prosody is often thought of as the music of language because it involves the modulation of one's voice with respect to pitch, rate, rhythm, duration, and stress (Frick, 1985). In so doing, information is provided about a speaker's emotion (e.g., happy versus sad) or intention (e.g., statement versus questions) and the prosodic cues enable people to recognize the beginning and end of sentences (Baum & Pell, 1999). Variation in pitch (or voice F_0) is one of the leading acoustic cues that help people distinguish both linguistic and emotional prosody (Wells & Peppe, 2003). Coincidentally, F_0 is transcribed with great fidelity at the level of the brainstem (e.g.,(Johnson, Nicol, & Kraus, 2005; Krishnan & Parkinson, 2000; Krishnan, Xu, Gandour, & Cariani, 2005; Krishnan, Xu, Gandour, & Cariani, 2004; Wong, Skoe, Russo, Dees, & Kraus, 2007; Xu, Krishnan, & Gandour, 2006) and can be controlled vocally (Bauer & Larson, 2003; Burnett, Freedland, Larson, & Hain, 1998; Chen, Liu, Xu, & Larson, 2007; Larson, Sun, & Hain, 2007; Liu & Larson, in press; Sivasankar, Bauer, Babu, & Larson, 2005; Titze, 1994). Thus, the controlled study of F_0 lends itself well to understanding one aspect of prosody.

Prosody Deficits in ASD

One of the hallmarks of the language impairment in ASD is a deficit with respect to both production and perception of prosody in speech. In the expressive realm, verbal children with ASD may speak with a monotone or expressionless quality; they may use scripted or echolalic

speech; they have an unusual pattern of intonation that is difficult to follow or speak too softly or loudly (McCann, Peppe, Gibbon, O'Hare, & Rutherford, 2007; Paul, Augustyn, Klin, & Volkmar, 2005; Peppe, McCann, Gibbon, O'Hare, & Rutherford, 2007; Rapin & Dunn, 2003; Shriberg et al., 2001). Children with ASD are often reported to have trouble "reading" emotion or intention in other people (Korpilahti et al., 2006; Paul, Augustyn et al., 2005; Peppe et al., 2007). Having atypical prosody in speech (e.g., monotonicity) often makes it difficult to understand what individuals with ASD mean and being unable to interpret these acoustic cues in the speech of others may impede social communication with peers (McCann et al., 2007; Paul, Shriberg et al., 2005; Rutter, 1974).

Speech and Speech in Noise

The acoustics of speech have been described via a source-filter model (Fant, 1960) (Ladefoged, 2001; Titze, 1994). The natural shape of an individual's vocal apparatus plays a large role in how sounds are produced. For example, the vibrations of vocal folds represent the sound source and the lowest frequency, or rate, of vibration constitutes the F_0. Prosodic information (e.g., pitch and intonation) is conveyed in part through adjustments in the rate of vocal fold vibrations. The modulation of air flow through vocal filters (e.g., vocal tract, oral cavity, tongue, lips and jaw) enables people to form different sounds. Manipulations of the filter cause the vocal tract to resonate and the resulting resonant frequencies are known as formants. Linguistic information about consonants and vowels (e.g., onsets and offsets) is conveyed largely through filter manipulations.

Even in normal and non-spectrum populations, decoding speech in background noise is problematic at best (Bradlow, Kraus, & Hayes, 2003; Picard & Bradley, 2001; Klatte, Meis, Sukowski, & Schick, 2007). Speech signals have both source (F_0-related) and filter aspects (onsets, offsets, and higher frequency components). In background noise, one can usually still distinguish the source-related aspects and recognize who a speaker is, even if s/he cannot differentiate what is being said. Further, vowels and periodic information are usually preserved in background noise, whereas consonants and rapid transitions get masked.

Children with ASD often report a hyper-sensitivity to sound and an increased perception of loudness, regardless if it is a relatively quiet or noisy environment (Gomes, Rotta, Pedroso, Sleifer, & Danesi, 2004; Khalfa et al., 2004; Minshew & Hobson, 2008; Rosenhall, Nordin, Sandstrom, Ahlsen, & Gillberg, 1999). However, often times this hyper-sensitivity is reported amidst normal hearing thresholds. There is relatively little physical evidence to explain this phenomenon. Individuals with ASD demonstrated poorer sound localization in background noise (Teder-Salejarvi, Pierce, Courchesne, & Hillyard, 2005) and evidence for broader auditory filter widths in high-functioning children with autism which may contribute to poor frequency discrimination in speech (Plaisted, Saksida, Alcantara, & Weisblatt, 2003). In a recent controlled study of speech perception in background noise, researchers identified elevated speech perception thresholds in noise in subjects with ASD (Alcantara, Weisblatt, Moore, & Bolton, 2004). Given that the neural processing of speech in noise at either the brainstem or cortical level is undefined in this population, one of the goals of this study was to characterize transcription of speech in noise at the brainstem and cortex.

Peripheral and Central Auditory Processing Deficits

When a child presents with absent or delayed language, it is common practice to test his or her hearing to rule out deafness or other peripheral deficits as the cause. Not surprisingly, the results of peripheral auditory processing measures in ASD are mixed; only some children present with peripheral hearing deficits (Rosenhall et al., 1999) (Gravel, Dunn, Lee, & Ellis, 2006; Khalfa et al., 2001). However, in addition to disrupted peripheral representations of sound, faulty neural representations of sounds can lead to a misinterpretation of auditory cues and essentially, miscommunication. Specifically, normal perception of speech depends on accurate timing of brainstem neurons (Johnson, Nicol, Zecker, & Kraus, 2007; Kraus & Nicol, 2005). Although the click-evoked auditory brainstem response has been used clinically to assess hearing in children with ASD because it is non-invasive and independent of state of the individual (i.e., awake or asleep), this response is limited in its ability to assess the integrity of the brainstem. As such, reports of click-evoked auditory brainstem responses are inconsistent (Klin, 1993; Rosenhall, Nordin, Brantberg, & Gillberg, 2003; Tanguay, Edwards, Buchwald, Schwafel, & Allen, 1982), suggesting that some children with ASD present with normal click-evoked responses (Courchesne, Courchesne, Hicks, & Lincoln, 1985; Rumsey, Grimes, Pikus, Duara, & Ismond, 1984; Seri, Cerquiglini, Pisani, & Curatolo, 1999; Tharpe et al., 2006) and others demonstrate abnormal responses (Maziade et al., 2000; McClelland, Eyre, Watson, Calvert, & Sherrard, 1992).

In order to uncover the underlying deficit in the language and prosody impairment, more environmentally salient stimuli (i.e., speech) would prove a more useful tool. The brainstem

transcribes source and filter aspects of speech with remarkable fidelity (Johnson et al., 2007; Kraus & Nicol, 2005). Unlike click-evoked responses, the speech-evoked response includes a "frequency-following response" (FFR) that mimics the frequency composition of speech sounds and phase locks to the periodic components (Akhoun et al., in press; Galbraith, Amaya et al., 2004; Hoormann, Falkenstein, Hohnsbein, & Blanke, 1992; Johnson et al., 2007; Russo, Nicol, Musacchia, & Kraus, 2004). A speech-evoked brainstem paradigm can also be used to describe how the brainstem encodes variations in F_0 (pitch contour) (Krishnan et al., 2005; Krishnan et al., 2004; Wong et al., 2007). Further, there is considerable evidence for a relationship between normal click-evoked responses and abnormal speech-evoked responses in children with other language-based learning disorders (Banai, Nicol, Zecker, & Kraus, 2005; Cunningham, Nicol, Zecker, Bradlow, & Kraus, 2001; Johnson et al., 2007; King, Warrier, Hayes, & Kraus, 2002; Wible, Nicol, & Kraus, 2004, 2005). Even so, the study of the relationship between speech-evoked brainstem deficits and language impairment in children with ASD is yet uncharted.

To date, the only studies investigating speech encoding as a basis for language problems in ASD have been restricted to the cortex. Long-latency (50-380 ms) responses to speech were prolonged and reduced, or non-existent, in children and adults who are on the autism spectrum (Ceponiene et al., 2003; Jansson-Verkasalo et al., 2003; Korpilahti et al., 2006; Kujala, Lepisto, Nieminen-von Wendt, Naatanen, & Naatanen, 2005; Lepisto et al., 2005; Lepisto et al., 2006). Further, dissimilar patterns of cortical activation (Boddaert et al., 2003; Boddaert, Chabane, Belin et al., 2004; Kasai et al., 2005; Oram Cardy, Flagg, Roberts, Brian, & Roberts, 2005) or atypical hemispheric lateralization to speech was also demonstrated in the ASD group (Boddaert

et al., 2003; Boddaert, Chabane, Gervais et al., 2004; Jansson-Verkasalo et al., 2003; Lepisto et al., 2006).

Finally, the effects of background noise on the neural processing of speech at either the brainstem or cortical level is uninvestigated in ASD. Given the reported perceptual difficulty extracting information from background noise in individuals with ASD (Alcantara et al., 2004; Plaisted et al., 2003; Teder-Salejarvi et al., 2005), understanding the neural transcription of speech in noise is a riveting issue to be studied in ASD.

In general, the benefit of recording auditory evoked potentials for characterizing deficits in ASD is that one can learn about temporal and spectral deficits in neural processing. With the accuracy of timing and frequency transcription by the auditory brainstem, it is possible to extract information about which specific aspects of speech encoding are disrupted (e.g., speech onset versus pitch). Further, brainstem response morphology is well suited for describing abnormal processing within individuals, rather than just at a group level. Beyond the brainstem, recording cortical evoked potentials in response to speech in quiet and background noise in the same children provides information about whether deficits are pervasive in the auditory system or whether distinct subgroups of children will show different patterns of normal or deficient activity.

Audio-vocal Interactions

Evidence from literature on birdsong development (Doupe & Kuhl, 1999; Margoliash, 2002; Prather, Peters, Nowicki, & Mooney, 2008) and speech production in cochlear implant patients

(Campisi et al., 2005; Hamzavi, Deutsch, Baumgartner, Bigenzahn, & Gstoettner, 2000; Higgins, McCleary, & Schulte, 1999; H. Lane et al., 1997; Leder et al., 1987; Monini, Banci, Barbara, Argiro, & Filipo, 1997; Perkell, Lane, Svirsky, & Webster, 1992; Svirsky, Lane, Perkell, & Wozniak, 1992) demonstrates that auditory feedback and spontaneous experimentation with language is crucial in learning to modulate vocal sounds. Perceptual exposure to sounds and auditory feedback aid in vocal production (Guenther, 2006; Guenther, Ghosh, & Tourville, 2006; Guenther, Hampson, & Johnson, 1998; Hain et al., 2000; Rutter, 1974; Smotherman, 2007; Tourville, Reilly, & Guenther, 2007). Thus, the proper transcription of sound is important in the development of appropriate patterns in language and to differentiate between sounds. A pitch-shift reflex paradigm allows the objective characterization of the audio-vocal system and, more specifically, the relationship between auditory feedback and control of voice F_0 (Burnett et al., 1998; Hain et al., 2000). This technique involves the vocalization of a steady-state voice F_0in the presence of conflicting auditory pitch-shifted feedback. The resulting response (corrections for errors in F_0) is a measure of the reflexive mechanism for regulating voice F_0 based on auditory feedback.

As mentioned previously, children with ASD have abnormal early experiences with language and the production of sounds (Dawson et al., 2000; Gernsbacher, 2004; Iverson & Wozniak, 2007; Rutter, 1974). Additionally, given the likelihood of auditory deficits in individuals with ASD, it is hypothesized that children with ASD who have abnormal prosody in speech may exhibit deficits in audio-vocal regulation. Because manipulations in voice F_0 comprise a main

aspect of prosody in speech, the pitch-shift reflex paradigm is useful in describing any audio-vocal system deficits in the automatic regulation of F_0 in ASD.

Corticofugal Interactions

The corticofugal pathway consists of descending fibers from the auditory cortex to lower levels. Such fibers are capable of shaping sensory events at the brainstem. Additionally, there are reciprocal pathways between the auditory brainstem and cortex (Ahissar & Hochstein, 2004; Eliades & Wang, 2003; Galbraith, Gutterson et al., 2004; Kraus & Banai, 2007; Suga, Gao, Zhang, Ma, & Olsen, 2000; Yan & Suga, 1996). Both bottom-up and top-down processing deficits can affect language development; that is to say that deficient initial sensory input to the cortex by the brainstem or poor cortical modulation of incoming sounds transcription by the brainstem may disrupt language. Further, modulation of cortical excitation and inhibition also affects the use of auditory feedback in production of vocalizations (Eliades & Wang, 2003). Long-range (long distance) functional connectivity between the brainstem and the cortex is necessary for the successful processing of information. Synapses become strengthened with persistent activation and become inactive after periods of no use.

Many different studies, invoking several experimental paradigms (e.g., PET, MEG, MRI, DTI), show evidence of reduced and aberrant long-range connectivity in ASD (Baron-Cohen, Knickmeyer, & Belmonte, 2005; Courchesne & Pierce, 2005; Just, Cherkassky, Keller, & Minshew, 2004; Minshew & Williams, 2007; Wickelgren, 2005). Presumably, faulty corticofugal pathways are involved in the speech-related neural processing deficits and language

impairment in ASD. The constellation of the aforementioned brainstem, cortical, and auditory-motor experimental paradigms together lend insight to this discussion of disrupted corticofugal pathways in ASD.

Scientific Contribution

At a basic science level, this work contributes to the understanding of the brainstem representation of speech and pitch contour, characterizes audio-vocal regulation in children and adolescents with and without ASD, and provides the first account of cortical encoding of speech in noise in children with ASD. Each of these accomplishments represents a substantial contribution to the fields of neuroscience, communication disorders, and linguistics; thus, allowing fellow scientists to further their respective research in such areas. Finally, the outcomes of these studies will broaden the understanding of sensory processing deficits and their relationship to the language impairment in ASD.

CHAPTER II: BRAINSTEM RESPONSES TO SPEECH SYLLABLES

Published in *Clinical Neurophysiology* 2004 Sep;115(9):2021-30.

Nicole Russo[a,b,c,*], Trent Nicol[a,c], Gabriella Musacchia[a,c], Nina Kraus[a,b,c,d,e]

[a]Auditory Neuroscience Laboratory, Northwestern University, 2240 Campus Drive, Evanston, IL 60208, USA [b]Institute for Neuroscience, Northwestern University, Evanston, IL 60208, USA [c]Department of Communication Sciences, Northwestern University, Evanston, IL 60208, USA [d]Department of Neurophysiology and Physiology, Northwestern University, Evanston, IL 60208, USA [e]Department of Otolaryngology, Northwestern University, Evanston, IL 60208, USA

Abstract

Objective: To establish reliable procedures and normative values to quantify brainstem encoding of speech sounds.

Methods: Auditory brainstem responses to speech syllables presented in quiet and in background noise were obtained from 38 normal children. Brainstem responses consist of transient and sustained, periodic components—much like the speech signal itself. Transient peak responses were analyzed with measures of latency, amplitude, area, and slope. Magnitude of sustained, periodic frequency-following responses was assessed with root mean square, fundamental frequency, and first formant amplitudes; timing was assessed by stimulus-to-response and quiet-to-noise inter-response correlations.

Results: Measures of transient and sustained components of the brainstem response to speech syllables were reliably obtained with high test–retest stability and low variability across subjects. All components of the brainstem response were robust in quiet. Background noise disrupted the

transient responses whereas the sustained response was more resistant to the deleterious effects of noise.

Conclusions: The speech-evoked brainstem response faithfully reflects many acoustic properties of the speech signal. Procedures to quantitatively describe it have been developed.

Significance: Accurate and precise manifestation of stimulus timing at the auditory brainstem is a hallmark of the normal perceptual system. The brainstem response to speech sounds provides a mechanism for understanding the neural bases of normal and deficient attention-independent auditory function.

Keywords: Speech syllable response; Brainstem response; Auditory brainstem response; Frequency-following response; Effects of noise

Introduction

The neural encoding of sound begins in the auditory nerve and travels to the auditory brainstem. Brainstem responses to simple stimuli (e.g., clicks, tones) are widely used in clinical practice in the evaluation of auditory pathway integrity (Moller & Jannetta, 1985; Starr & Don, 1988). Less well-defined is how the brainstem responds to complex stimuli. Describing auditory encoding of speech sounds provides insight into some of the central auditory processes involved in normal communication. Furthermore, this knowledge may be applied to understanding effects of the aging process on hearing, as well as to a broad range of other circumstances, including hearing

and communication in individuals with learning problems, peripheral hearing impairments, cochlear implants, or auditory neuropathies.

Background and Significance

Some people have normal peripheral hearing, but still cannot perceive speech well. Previous studies have shown that the disruption of neural timing at the cortex is linked to auditory perceptual deficits (Kraus et al., 1996; Nagarajan et al., 1999) (Tonnquist-Uhlen, 1996; Wible, Nicol, & Kraus, 2002). In addition, abnormal electrophysiological responses to speech syllables at the brainstem level have been associated with a wide spectrum of diagnosed learning problems (King et al., 2002; Wible et al., 2004). These abnormalities include a temporally delayed response to the onset of a consonant and deficient spectral representation of harmonic aspects of the speech signal. Disruptions of neural encoding in both the brainstem and cortex were exacerbated when speech was presented in background noise (Cunningham et al., 2001).

Part of the difficulty in perceiving consonants in noisy situations is that they are rapid, relatively low-amplitude transient features of speech. Stop consonants, such as /d/, are known to be particularly vulnerable to disruption by background noise in normal and clinical populations (Brandt & Rosen, 1980). The perception of vowels, however, is more resistant to the effects of noise because they are periodic, sustained signals, and generally louder than consonants.

Brainstem responses provide direct information about how the sound structure of a speech syllable is encoded by the auditory system. It is particularly compelling to consider that specific

aspects of the sound structure of the acoustic signal are maintained and reflected in the neural code. Similar to the speech syllable itself, the brainstem response to a speech syllable can be divided into transient and sustained portions, namely the onset response and the frequency-following response (FFR) (Boston & Moller, 1985). Onset responses are transient, with peak durations lasting tenths of milliseconds, thus we will refer to these rapid deflections as transient responses. Within the FFR are discrete peaks corresponding to the periodic peaks in the stimulus waveform. However, this region can be considered as a whole, as it contains a periodic signal sustained for tens or hundreds of milliseconds. Although peaks within the FFR may be thought of as successive onsets, for descriptive purposes, we will use the term FFR to refer to the later portion of the response evoked by the harmonic vowel structure of the stimulus. There is a parallel effect of noise on the brainstem response, similar to the disruption of speech perception, in that transient onsets were more affected by the noise, sometimes even eliminated, while the sustained portion remained intact (Cunningham et al., 2001).

Specific Aims

The specific aims of this study were: (1) to delineate measures of the timing and magnitude of the brainstem response to the speech syllable /da/ in quiet and background noise; (2) to establish normative values for these features; and (3) to determine the test–retest reliability of these measures.

Methods

Subjects

Thirty-eight children, ages 8–12 years (21 male, 17 female) participated in the primary focus of this study, which established normative values for the brainstem response to speech syllables. Eight children (four male, four female) were part of the retest reliability portion of the study. None of the children had a history of medical or learning problems and all performed within normal limits on laboratory-internal standardized measures of learning and academic achievement. These measures consisted of selected subtests of Woodcock Johnson, Woodcock Johnson—Revised, and Wide Range Achievement Tests that have been described in detail elsewhere (Kraus et al., 1996). All of the subjects had normal click-evoked auditory brainstem response latencies and normal hearing thresholds at or below 20 dB HL for octaves from 500 to 4000 Hz. Consent and assent were obtained from the parents (or legal guardians) and the children involved in the study. The Institutional Review Board of Northwestern University approved all research.

Stimulus and Recording Parameters

Because stop consonants provide considerable phonetic information and their perception is particularly vulnerable to background noise in both normal and clinical populations, a five-formant synthesized /da/ was chosen for the stimulus (Klatt, 1980). The stimulus duration was 40 milliseconds (ms). The consonant contained an initial 10 ms burst; the frequencies of which were centered around the beginning frequencies of formants 3–5, thus in the range of 2580–4500 Hz. Additional details of the speech synthesis parameters can be found in King et al. (2002). The syllable /da/ was presented monaurally, in alternating polarities, at 80 dB SPL to the right ear via

insert earphones (ER-3, Etymotic Research, Elk Grove Village, IL), with an inter-stimulus interval of 51 ms.

During testing, children watched a videotape with the sound level set at , 40 dB SPL in the non-test ear. Responses were collected with Ag–AgCl scalp electrodes, differentially recorded from Cz (active) to ipsilateral earlobe (reference), with the forehead as ground. Three blocks of 1000 sweeps per block were collected at each polarity and in each of two different signal-to-noise conditions, quiet and ipsilateral white Gaussian noise (þ5 dB SNR). Waveforms were averaged online in Neuroscan (Compumedics, El Paso, TX) with a recording time window spanning 10 ms prior to the onset and 20 ms after the offset of the stimulus. The sampling rate was 20,000 Hz and responses were online bandpass filtered from 100–2000 Hz, 6 dB/octave. Trials with eye-blinks greater than 35 mV were rejected online. Responses of alternating polarity were then added together to isolate the neural response by minimizing stimulus artifact and cochlear microphonic (Gorga, Abbas, & Worthington, 1985).

Description of the Brainstem Response to Speech Syllables

The electrophysiological brainstem response to a speech sound is a complex waveform (Fig. 1). This waveform includes transient peaks as well as sustained elements that comprise the FFR. The response to the onset of the speech stimulus /da/ includes a positive peak (wave V), likely analogous to the wave V elicited by click stimuli, followed immediately by a negative trough (wave A). In most subjects, positive peaks corresponding to click-evoked auditory brainstem response waves I and III are also visible. Following the onset response, peaks C and F are

present in the FFR. While other peaks are discernable in this region, a previous study (King et al., 2002) determined that peaks C and F were the most stable for this stimulus across individuals, having latencies with standard deviations (SDs) less than 0.5 ms across a normal population. The defining feature of the sustained portion of the response is its periodicity, which follows the frequency information contained in the stimulus (Marsh et al., 1974; Smith et al., 1975). The timing and magnitude of both the transient and sustained aspects of the response waveform were evaluated with the measures described below and summarized in Table 1.

Discrete Peak Measures

Measures of both timing and magnitude were utilized to assess the discrete peaks. The onset response waves V and A occur at latencies before 10 ms, while peaks C and F occur at latencies of about 18 and 40 ms, respectively. Three experienced raters picked peaks V, A, C, and F and their latencies and amplitudes were measured. The VA complex was further investigated by measuring its inter-peak interval, amplitude, slope, and area. These measures were taken from the raw waveforms of the responses collected in quiet. When background noise was introduced with the syllable, peaks were often obscured in the raw waveform. Therefore, a wavelet-denoising technique adapted from Quian Quiroga and Garcia (Quian Quiroga & Garcia, 2003) was applied to the waveforms collected in noise before obtaining transient response measures. Nevertheless, some peaks were still imperceptible after the denoising procedure. These peaks were designated as having 0 mV for amplitude and area and were omitted from latency, slope, and inter-peak interval analyses. Additionally, some peaks were not eliminated completely, but if there was not a consensus among peak pickers regarding the actual presence and location of the

peak, it was omitted from statistical analyses. These omissions are reflected in the subject numbers listed in Tables 2–4.

Frequency-following Response Measures

The region following the onset response was defined as the FFR. The FFR analysis window was chosen to incorporate as much of the sustained response as possible, while avoiding the refractory period of the onset response and any offset transients. Five analysis techniques were employed to analyze the FFR: (1) root mean square amplitude (RMS amp); (2) amplitude of the spectral component corresponding to the stimulus fundamental frequency (F0 amp); (3) amplitude of the spectral component corresponding to first formant frequencies of the stimulus (F1 amp); (4) stimulus-to-response (S–R) correlations, and (5) inter-response (I–R) correlations between the responses collected in the quiet and noise conditions. These composite FFR measures describe the sustained portion of the response as whole.

Root mean square amplitude

This measure reflects the averaged magnitude of activation of the neural population over an 11.5–46.5 ms epoch of the sustained response. Responses were de-meaned and, to correct for varying amounts of internal (e.g., myogenic) noise among subjects, the RMS amplitude of the response was divided by the RMS amplitude of the pre-stimulus period.

Amplitudes of the fundamental frequency and first formant

The FFR consists of energy at the fundamental frequency of the stimulus and its harmonics (Worden & Marsh, 1968). Fourier analysis was performed on the 11.5–46.5 ms epoch of the FFR in order to assess the amount of activity occurring over two frequency ranges. Activity occurring in the frequency range of the response corresponding to the fundamental frequency of the speech stimulus (103 –121 Hz) was calculated for each subject. The response activity corresponding to the first formant frequencies of the stimulus (220–720 Hz) was also measured. The F0 amp provides a gauge of the specific portion of the sustained response devoted to encoding the fundamental frequency of the speech sound, while the F1 amp is devoted to encoding the first formant (Fig. 2). A 2 ms on–2 ms off Hanning ramp was applied to the waveform. Zero-padding was employed to increase the number of frequency points where spectral estimates were obtained. A subject's response was required to be above the noise floor in order to be included in the analyses. This calculation was performed by comparing the spectral magnitude of the pre-stimulus period to that of the response[1]. If the quotient of the magnitude of the F0 or F1 frequency component of the FFR divided by that of the pre-stimulus period was greater than or equal to one, the response was deemed above the noise floor. The raw amplitude value of the F0 or F1 frequency component of the response FFR was then measured. Only the F0

[1] The FFR period is 3.5 times longer than the pre-stimulus period, so in order to make an accurate comparison, the spectral magnitude of the 10 ms pre-stimulus period was compared against the average of three 10 ms ranges (12.5–22.5, 22.5–32.5, and 32.5–42.5 ms) within the FFR. This ensured that the same number of points and therefore the same frequency ranges were compared. The pre-stimulus period and each of the 10 ms ranges were de-meaned to zero before performing spectral analyses.

and F1 frequencies of the response were above the noise floor. Response frequencies corresponding to higher stimulus formants were not significantly above the noise floor in either quiet or background noise.

Stimulus-to-response correlation

The stimulus-to-response (S–R) correlation reflects how faithfully the response waveform mimics the stimulus waveform, and provides a measure of phase locking that excludes the non-periodic activity inherent in the RMS amp measure. Each response was cross-correlated to the 10–40 ms portion of the stimulus that includes the harmonic segment of the syllable. Due to the time it takes for neural impulses to propagate through the nervous system, the response lagged behind the stimulus by approximately 7–10 ms. Thus, maximal correlation within this range was recorded.

Quiet-to-noise inter-response correlation

The inter-response (I–R) correlation reflects the fidelity of the response morphology recorded in noise to that of the response recorded in quiet, providing a way to quantify the effects of background noise on the timing of the sustained response. The I–R correlation was calculated similarly to the S–R correlation measure. However, because the addition of noise can delay the brainstem response by a couple of milliseconds, a cross-correlation was performed and the quiet response was allowed to lead the noise response by up to 2 ms.

Mean Pearson's r-values were reported for normative descriptive purposes of S–R and I–R correlations, although Fisher's transformation was used to convert r-values to z0-scores for all statistical computations.

Results

Based on our evaluation of 38 subjects' responses recorded in quiet and 36 subjects' responses recorded in background noise, normative values for the aforementioned brainstem measures were established. Table 2 shows means and SDs for discrete peak measures obtained in quiet and background noise. Tables 3 and 4 provide timing and magnitude values, respectively, for the FFR.

The Normal Response in Quiet

Responses were highly replicable both within and across subjects. Fig. 3 shows three 1000-sweep blocks obtained from a representative subject (top), as well as responses obtained from another subject on two separate test sessions (bottom). Peaks V, A, and C were detectable in all subjects, and peak F was detectable in all but two subjects. The onset response waves V and A were largest in magnitude, followed by FFR peak F and then peak C. Consistent with other neurophysiological responses, as latency increases, so does its variability. The SD of latency was smallest for the early onset response waves V and A (0.25 and 0.34 ms, respectively), and increased with latency (up to 0.61 ms).

The FFR was evident in all subjects. Timing of the FFR was indicated by S–R and I–R correlations. The magnitude of the response was evaluated with RMS, F0 and F1 amplitudes. As shown in Fig. 2, the greatest amount of energy is present in the F0 region.

Relationships among brainstem response measures

Relationships among brainstem response measures recorded in quiet were explored using Pearson's correlations. Correlations among measures exceeding ^0.30 and P , 0:05 criterion were considered significant. Tables 5–7 show the relationships among and between transient and sustained measures. Transient measures, especially those describing the VA complex and wave C, were largely related to each other (Table 5), while the timing of peak F was relatively independent of the timing of other peaks. Composite sustained measures did not exhibit strong relationships with each other, indicating that each measure described a unique quality of the FFR (Table 6). Transient onset responses were largely independent of sustained measures, with the exception that a number of wave V and A measures were related to F1 amplitude. Discrete peaks within the FFR were related to almost all of the composite FFR measures (Table 7).

The Normal Response in Noise

The addition of background noise interfered with normal brainstem encoding of the speech stimulus /da/. Fig. 4 shows the effects of noise. Table 2B shows normative values for the transient response measures in noise. Most affected were the onset responses V and A, which were severely degraded and completely obscured in more than 40% of the subjects. Peaks C and F, however, remained present in noise in most subjects (100 and 86%, respectively). Peak

amplitudes also were affected by the presence of noise; all peaks were reduced in size (P<0:001; all tests).

When not eliminated, latencies of onset peaks V and A, and FFR peak C were delayed in comparison to quiet (p<0.01; all tests). In contrast, peak F showed resilience to background noise in that its latency did not change with the addition of the noise (P<0.12) and remained easily identifiable in most subjects. As expected, the introduction of background noise increased the variability in the latencies of all peaks.

Although reduced, the composite FFR remained relatively intact and was discernible in noise. RMS amp and S–R correlations showed significant reductions in noise (p<0:002; both tests). F0 and F1 amp were also significantly affected by the presence of background noise (p<0:002; both tests). The addition of noise obscured onset peaks in the responses of many subjects, thus it was not possible to calculate the relationships between onset and FFR measures in noise.

Test–retest Stability

In order to determine whether the variables described here are stable over time, eight of the children were retested after a 2–10-month interval. Test–retest reliability is illustrated in the waveforms shown in a representative subject in Fig. 3 (bottom) and at the group level in Fig. 4. Two-tailed, paired t tests were calculated for all brainstem response measures. A significance criterion of (p<0.05) was used. Most brainstem measures did not change significantly over the test–retest time interval (p>0.09); exceptions included the amplitude and slope of the VA

complex in quiet and wave C latency in noise ($p<0.02$; all). Sustained measures were stable from test to retest ($p>0.30$; all tests).

Discussion

The ability to quantify a brainstem response elicited by speech sounds provides a powerful tool for research and clinical use. The speech-evoked brainstem response faithfully reflects many acoustic properties of the speech signal. In the normally perceiving auditory system, stimulus timing, on the order of fractions of milliseconds, is accurately and precisely represented at the level of the brainstem. Overall, the brainstem response provides a mechanism for understanding the neural bases of normal and deficient auditory function, by providing a quantifiable measure of an individual's attention-independent neural encoding of speech sounds.

This study described explicit methods to record and quantify the brainstem response to /da/ in quiet and in background noise and provided a normative data set which can be used to assess the integrity of speech signal encoding in normal and clinical populations. Measures of timing and magnitude were identified for both transient and sustained aspects of the responses. Transient response measures included latency and amplitudes of peaks V, A, C, and F, as well as inter-peak interval, slope, area and amplitude of the VA complex as a unit. Sustained measures included RMS amplitude, F0 and F1 amplitudes, S–R correlations, and I–R correlations. In quiet, these brainstem encoding parameters can be obtained nearly 100% of the time; variability is low and test–retest stability is high. The addition of background noise often eliminated the onset response (waves V and A) or resulted in non-uniform latency delays. Because robust responses

are necessary for accurate encoding, this disruption could underlie perceptual difficulties. Although the latencies of waves V, A, and C were delayed in noise, peak F remained stable. Thus, while it appeared that background noise induced a delay in responding to the onset of a sound, compensatory mechanisms may correct for this lag throughout the neural pathway. F0 remained robust in background noise, while other sustained measures, despite often being reduced in magnitude, also showed more resilience to the effects of noise. Overall, test–retest stability was high for responses obtained in both quiet and background noise. Although minimal variability may exist due to placement of electrodes or the insert earphone, the test–retest measures described in this study showed considerable stability over time.

Interpreting the Brainstem Response: Transient Versus Sustained

In as much as it may be an oversimplification to equate features of speech, such as consonants and vowels, with transient and sustained evoked responses, there are certain parallels. The transient portions of the brainstem response reflect the encoding of rapid temporal changes inherent in consonants. The sustained FFR encodes the harmonic and periodic sound structure of vowels. In quiet conditions, both the transient and sustained components of the speech syllable /da/ are robustly encoded. In noise, just as vowels are less affected than consonants, the FFR is less degraded than the onset response.

A major difference between the onset and FFRs measured here was that under a stressed circumstance— background noise—neural encoding of onset features was severely degraded, whereas the sustained FFR features remained relatively unaltered. Onset waves V and A were

eliminated in almost half of the subjects, while peaks C and F, and the FFR region as a whole, remained stable. Consequently, the perceptual problems associated with consonant identification in background noise could be attributed to the decreased neural synchrony reflected in the onset, while the intact encoding reflected in the sustained region enabled accurate vowel perception. F0 amplitude remained robust in noise. Encoding of the fundamental frequency is important for identifying the speaker and emotional tone of voice. Meanwhile, the degradation of F1; which provides phonetic information, coupled with the loss of the onset burst, further degrades perception of the speech signal in noise. These data provide evidence to support the observation that speaker identity and speaker tone of voice is more resistant to noise than the phonetic content of what is being said.

However, another possible explanation is that the /da/ stimulus is smaller in amplitude at its onset than towards the end. Thus, the elimination of waves V and A, and the maintenance of the FFR, may be due to the relative differences in which aspects of the stimulus did or did not exceed the level of the acoustic masking noise. Future studies incorporating different types of background noise, such as pink noise or multi-speaker babble maskers (which more closely resemble naturally occurring noise and the spectrum of speech itself) likely will contribute to further understanding the encoding of speech in background noise.

The overall resistance of the FFR versus the disruption of the onset response in noise suggests a relative independence of brainstem encoding processes. Furthermore, the independence of the transient versus sustained responses was apparent in the relationships among measures. That is,

correlations were strong within transient and sustained measures separately, whereas fewer, weaker relationships existed between these classes of measures. Although transient measures within the FFR (e.g., waves C and F) showed relationships to the composite sustained response measures, transient onset and composite FFR measures demonstrated few relationships, reinforcing that they are neither wholly separate nor wholly related measures. It is interesting to note that brainstem responses that reflect prosodic aspects of speech (F0 and RMS amplitude) are largely independent from the internally related measures (waves V, A, and C latency and F1 amplitude), which represent phonetic information of the stimulus.

Practical Applications

Individual versus group data

Most physiological and imaging approaches for assessing the functional integrity of sensory systems require group data and can be time-intensive. Collecting the brainstem response to a speech sound can be done in a few minutes, requires few electrodes, and is passively acquired. Normal variability of response attributes is low. Furthermore, the brainstem response is stable over time. Consequently, the measures reported here lend themselves to the assessment of the encoding of sound structure in individual subjects.

Identification of auditory-based learning disabilities

The data provided here serve as a metric for determining normal brainstem function in response to speech sounds. Deficits in neural timing and magnitude in response to speech syllables at the brainstem level have been previously found to occur in certain children with auditory-based

learning problems (Cunningham et al., 2001; King et al., 2002). Timing abnormalities in waves V, A, and C have been identified (King et al., 2002). Decreases in S–R correlations and the reduced magnitude of the FFR, specifically in the frequency composition of F1; have also been found in children with learning problems (Cunningham et al., 2001).

Predictors of future language impairment

A recent publication by Benasich and Tallal (Benasich & Tallal, 2002) reported that behavioral measures of central auditory function, obtained in children under 1 year of age (mean age=7.5 months), can serve as predictors for subsequent specific language impairments (SLI) and other developmental language delays. Due to the early maturation of the brainstem response, the brainstem measures described in this paper, might provide a biological marker for early detection of central auditory deficits that may dovetail with these behavioral findings. Further research is needed to determine which specific manifestations of brainstem abnormalities may facilitate the early prediction of SLI. The normative data provided here can serve as an objective index for early diagnosis and identification of deficits in the neural encoding of sound structure in the brainstem. Intervention could be applied before the behavioral aspects of their impairment impact a child.

Predictors for success with auditory training

Neural encoding of sound structure in the auditory brainstem may provide a predictive index for success with auditory training regimens. Children with learning problems and brainstem-encoding deficits have been shown to benefit from auditory training (Hayes, Warrier, Nicol,

Zecker, & Kraus, 2003). Specifically, trained children with a delayed brainstem onset latency (wave A) in quiet showed greater improvements in the timing and magnitude of cortical responses. Additionally, behavioral improvements were seen in tests of phoneme decoding (Sound Blending and Auditory Processing) in these children. Thus, children with brainstem encoding deficits particularly appeared to benefit from auditory training. These data support the idea that early analysis of the brainstem response could predict which children would benefit from auditory training.

Conclusions

Brainstem response timing and magnitude measures provide reliable information about the neural encoding of speech sounds. This study outlined specific measures of brainstem function that may be used to characterize neural encoding of speech sounds for clinical and research applications. Transient and sustained measures provide information regarding auditory pathway encoding of brief and periodic aspects of the stimulus. Some of the data suggest that transient and sustained responses represent independent mechanisms. A better understanding of brainstem encoding may assist in early diagnosis and intervention of auditory disorders, as well as in measuring the success of training programs.

The current study is a springboard for further examination of brainstem activity to complex speech stimuli, as well as for identifying abnormalities in clinical populations such as aging, peripheral hearing impairment, cochlear implant, auditory neuropathy, and non-native listener populations, in which the assessment of auditory function is relevant. Future parametric studies

(e.g., of the effects of different speech stimuli, methods of presentation or types of background noise) may enhance the potential clinical use of the brainstem response to speech sounds.

Acknowledgements

The National Institute of Health NIDCD R01-01510 supported this research. We thank the children and their families who participated in this study. We also thank Steven Zecker, for statistical consultation; Erika Skoe, for the development of programs to analyze the data; and also members of the Kraus laboratory who tested subjects and gave their support throughout the study.

CHAPTER III: BRAINSTEM TRANSCRIPTION OF SPEECH IS DISRUPTED IN CHILDREN IN CHILDREN WITH AUTISM SPECTRUM DISORDERS

Submitted for Publication in *Developmental Science*

Nicole Russo[a,b], Trent Nicol[a], Barbara Trommer[b,c], Steve Zecker[a], Nina Kraus[a,b,d]

[a]The Roxelyn and Richard Pepper Department of Communication Sciences, Northwestern University, Evanston, IL, [b]Northwestern University Interdepartmental Neuroscience Program, Evanston, IL, [c]Departments of Pediatrics and Neurology, Feinberg School of Medicine, Northwestern University, Chicago, IL and Evanston Northwestern Healthcare, Evanston, IL, [d]Department of Neurobiology and Physiology and Department of Otolaryngology, Northwestern University, Evanston, IL.

Abstract

Language impairment is a hallmark of autism spectrum disorders (ASD). The origin of the deficit is poorly understood although deficiencies in auditory processing have been detected in both perception and cortical encoding of speech sounds. Little is known about the processing and transcription of speech sounds at earlier (brainstem) levels or about how background noise may impact this transcription process. Unlike cortical encoding of sounds, brainstem representation preserves stimulus features with a degree of fidelity that enables a direct link between acoustic components of the speech syllable (e.g., onsets) to specific aspects of neural encoding (e.g., waves V and A). We measured brainstem responses to the syllable /da/, in quiet and background noise, in children with and without ASD. Children with ASD exhibited deficits in both the neural synchrony (timing) and phase locking (frequency encoding) of speech sounds,

despite normal click-evoked brainstem responses. They also exhibited reduced magnitude and fidelity of speech-evoked responses and inordinate degradation of responses by background noise in comparison to typically developing controls. Neural synchrony in noise was significantly related to measures of core and receptive language ability. These data support the idea that abnormalities in the brainstem processing of speech contribute to the language impairment in ASD. Because it is both passively-elicited and malleable, the speech-evoked brainstem response may serve as a clinical tool to assess auditory processing as well as the effects of auditory training in the ASD population.

Keywords: auditory brainstem, autism spectrum disorder, speech, language, evoked potentials

Introduction

Autism spectrum disorders (ASD) is a cluster of disorders that includes autism, Asperger syndrome, and Pervasive Developmental Disorder Not Otherwise Specified (PDD-NOS) (Siegal & Blades, 2003; Tager-Flusberg & Caronna, 2007). Impairment in the social and communicative use of language is a hallmark of ASD. In severe cases, children with ASD are non-verbal. When speech is present, it is often slow to develop, echolalic, stereotypic, emotionless or excessively literal (Boucher, 2003; Shriberg et al., 2001; Siegal & Blades, 2003). An individual with ASD will usually not engage in typical reciprocal communication (Rapin & Dunn, 2003; Tager-Flusberg & Caronna, 2007). Both receptive and expressive deficits can occur and both have been attributed, at least in part, to abnormalities of auditory processing (Siegal & Blades, 2003). Further, children with ASD have particular difficulties processing speech in

background noise, as demonstrated by high speech perception thresholds and poor temporal resolution and frequency selectivity (Alcantara et al., 2004).

The Source-filter Model of Speech

Successful communication relies on being able to both produce and process speech sounds in a meaningful manner. The literature on speech production provides a useful dichotomy to describe the acoustics of speech (Fant, 1960) (Ladefoged, 2001; Titze, 1994), the source-filter model. In this model, the vibration of the vocal folds represents the sound source. The lowest frequency of a periodic signal, such as speech sounds, is known as the fundamental frequency (F_0) and it is the rate of the vocal fold vibrations. Everything else--vocal tract, oral cavity, tongue, lips and jaw--comprises the filter. The vocal tract resonates and the resulting resonant frequencies are known as formants. The formants are conveyed through manipulations of the filter and provide cues about onsets and offsets of sounds. Broadly speaking, linguistic content--vowels and consonants--is transmitted by particular filter shapes, whereas nonlinguistic information such as pitch and voice intonation, relies largely on characteristics of the source. Although source-filter cues are simultaneously conveyed in the acoustic stream of speech, remarkably, they can be readily transcribed as both discrete components and as a whole by the auditory brainstem (Johnson et al., 2007; Kraus & Nicol, 2005).

Transcription in the Auditory Brainstem

The brainstem response has the capacity to reveal auditory pathway deficits in a non-invasive and passive manner, which has engendered its long history of clinical use even in difficult-to-test

populations. Timing and periodicity of an evoking stimulus are preserved in this response, enabling it to reflect processing deficits that arise from the peripheral auditory system and ascending auditory pathway. Soundwaves propagate through the auditory nerve, lower brainstem structures including the cochlear nucleus, superior olivary complex, lateral lemniscus and inferior colliculus (Buchwald & Huang, 1975; Hood, 1998; Moller & Jannetta, 1985). Precision is such that timing delays on the order of fractions of milliseconds are diagnostically significant and the latency of responses lends insight into where in the pathway the anomalies occur (Hood, 1998; Jacobson, 1985).

Click stimuli, which are typically used to assess hearing, evoke short latency auditory brainstem responses (≤10 ms) that provide information limited to timing and amplitude. The brainstem frequency-following response (FFR), which can not be elicited by a click, phase locks to the periodic components of a stimulus (in the case of speech, the source information) (Galbraith, 1994; Galbraith, Amaya et al., 2004; Galbraith, Philippart, & Stephen, 1996; Worden & Marsh, 1968) and is thought to originate in the auditory midbrain lateral lemniscus (Galbraith, 1994) and inferior colliculus (J. C. Smith, Marsh, & Brown, 1975). In one study of the FFR to phrase-speech presented forward and backward, the far-field brainstem response was enhanced (increased signal-to-noise ratio) for forward speech (Galbraith, Amaya et al., 2004), suggesting that the auditory brainstem may respond preferentially to stimuli with environmental significance or with which people have greater experience.

The Speech-evoked Auditory Brainstem Response in Quiet and Background Noise

Speech is a complex stimulus that, unlike a click, has environmental relevance. The speech-evoked brainstem response lends itself to the extraction of information about encoding of syllable onset, offset and periodicity (pitch and formant spectra). The brainstem response to the syllable /da/ reflects both phonetic/filter (transient) and prosodic/source (periodic) acoustic features with remarkable fidelity (Johnson et al., 2005; King et al., 2002; Kraus & Nicol, 2005; Russo et al., 2004) (Figure 5). The timing of these neural responses relays information about neural synchrony. Specifically, four transient responses (waves V, A, C, and O) and phase locking in the range of the first formant (F_1) convey the filter aspects of the speech syllable. The response to the onset of the syllable is indicated by a wave V and its negative trough, wave A. Wave C represents the transition to the periodic, voiced portion of the stimulus that corresponds to the vowel. Wave O corresponds to stimulus offset. The F_0, which transmits pitch cues, is reflected in the time domain by FFR waves D, E and F and phase locking to that frequency. Source cues are conveyed by the F_0 in both the frequency and time domains. Moreover, the brainstem response to speech presented in background noise provides an index of auditory pathway function in challenging listening situations. Even in the normal system, transient onset synchrony and phase locking to frequencies in the range of F_1 are significantly diminished, whereas the encoding of F_0 remains robust.

Clinical Correlations and Utility of the Auditory Brainstem Response

Auditory brainstem function has been linked to language impairment (Banai et al., 2005; Cunningham et al., 2001; Johnson et al., 2007; King et al., 2002; Wible et al., 2004, 2005) and also to auditory expertise, such that speech-evoked brainstem responses have been shown to be

shaped and enhanced by lifelong linguistic (Krishnan et al., 2005; Krishnan et al., 2004; Xu et al., 2006) and musical experience (Musacchia, Sams, Skoe, & Kraus, 2007; Wong et al., 2007), possibly through corticofugal feedback to subcortical sensory circuitry (Ahissar & Hochstein, 2004; Kraus & Banai, 2007). Although prior studies have investigated cortical evoked responses to speech and their relationship to language in individuals with ASD (Boddaert et al., 2003; Boddaert, Chabane, Gervais et al., 2004; Ceponiene et al., 2003; Jansson-Verkasalo et al., 2003; Kasai et al., 2005; Kuhl, Coffey-Corina, Padden, & Dawson, 2005) (Lepisto et al., 2007; Lepisto et al., 2005; Lepisto et al., 2006), the majority of studies of the auditory brainstem have focused on responses to non-speech stimuli (i.e., clicks or pulses) (reviewed in (Klin, 1993; Rapin & Dunn, 2003). Such studies yielded mixed results, reporting abnormal, increased conduction times (Maziade et al., 2000; McCabe, Smith, & LePore, 2000), normal responses (Courchesne et al., 1985; Rumsey et al., 1984; Tharpe et al., 2006) and deficits in only a subset of children (Rosenhall et al., 2003). These varying results may reflect the diagnostic heterogeneity in ASD. In the only known study of speech-evoked brainstem responses in ASD, Russo and colleagues (Russo et al., in press) reported deficient processing of pitch cues in speech in a subset of children with ASD.

In the present study, we utilized the speech-evoked auditory brainstem response to investigate the language deficiencies characteristic of individuals with ASD. Our hypotheses were that 1) children with ASD would demonstrate deficits in auditory brainstem function relative to their typically-developing (TD) counterparts and 2) that abnormalities in the neural processing of speech in the ASD group would correlate with their language impairment. Children participating

in this study were required to have normal auditory brainstem responses to clicks, thereby ruling out middle ear or cochlear dysfunction(Hood, 1998; Jacobson, 1985; Rosenhall et al., 2003) and so that any intergroup differences would be attributable only to difficulties processing speech.

Methods

The Institutional Review Board of Northwestern University approved all research and consent and assent were obtained both from the parent(s) or legal guardian(s) and the child.

Children were acclimated to the testing location and equipment prior to experimental data being collected. They were allowed to visit the laboratory and interact with the tester on multiple occasions. Some children brought electrodes home with them to better familiarize themselves with the neurophysiological procedure.

Participants

Of the 45 children originally recruited for this study, six (all children with ASD) were excluded for the following reasons: abnormal click-evoked brainstem responses (N = 2), mental ability below inclusion cutoff (N = 1), non-compliance resulting in inability to test (N = 1), parental decision to discontinue due to the required time commitment of the study (N = 1) and relocation (N = 1). Participants included 21 verbal children with ASD (N =19 boys, 2 girls) and 18 typically-developing children (TD, N = 10 boys, 8 girls). Age range was 7-13 years old and

mean age (*SD*) did not differ between groups (TD, $M = 9.7$, $SD = 1.934$ vs. ASD, $M = 9.90$, $SD = 1.921$; independent two-tailed t-test; $t(37) = .295$, $p = .77$).

Study participants were recruited from community and internet-based organizations for families of children with ASD. Children in the diagnostic group were required to have a formal diagnosis of one of the ASDs made by a child neurologist or psychologist and to be actively monitored by their physicians and school professionals at regular intervals. Parents were asked to supply the names of the examining professionals, their credentials, office location, date of initial evaluation and the specific diagnosis made. Diagnoses included autism (n=1), Asperger Disorder (n=7), PDD-NOS (n=1), and a combined diagnosis (e.g., Asperger Disorder/PDD-NOS; n=12). The diagnosis of ASD was supplemented by observations during testing such that included subjects were noted to have some or all of the following: reduced eye contact, lack of social or emotional reciprocity; perseverative behavior; restricted range of interests in spontaneous and directed conversation; repetitive use of language or idiosyncratic language; abnormal pitch, volume, and intonation; echolalia or scripted speech; and stereotyped body and hand movements. Diagnosis was also supplemented by an internal questionnaire that provided developmental history, a description of current symptoms, and functional level at time of entry into the study.

Further inclusion criteria for both TD and ASD groups were 1) the absence of a confounding neurological diagnosis (e.g., active seizure disorder, cerebral palsy), 2) normal peripheral hearing as measured by air threshold pure-tone audiogram and click-evoked auditory brainstem

responses and 3) a full-scale mental ability score whose confidence interval included a value ≥ 80.

Hearing Screening

On the first day of testing, children underwent a hearing threshold audiogram for bilateral peripheral hearing (≤ 20 dB HL) for octaves between 250 and 8000 Hz via an air conduction threshold audiogram on a Grason Stadler model GSI 61. Children wore insert earphones in each ear and were instructed to press a response button every time they heard a beep. At each subsequent test session, follow-up hearing screenings at 20 dB HL for octaves between 125 and 4 kHz were conducted using a Beltone audiometer and headphones.

Cognitive and Academic Testing

All cognitive and academic testing took place in a quiet office with the child seated across a table from the test administrator. Full-scale mental ability was assessed by four subtests of the Wechsler Abbreviated Scale of Intelligence (WASI, four subtests) (Woerner & Overstreet, 1999). This testing was supplemented by scores of performance (Block Design and Matrix Reasoning subtests) and verbal (Vocabulary and Similarities) mental ability which were not part of inclusion criteria (Table 8; mean and standard deviations). The Clinical Evaluation of Language Fundamentals-4 (CELF) (Semel, Wiig, & Secord, 2003) was administered to provide indices of core, expressive and receptive language abilities (Table 8). However, performance on the CELF was not used as a study inclusion criterion.

Stimuli and Data Collection

All neurophysiological recordings took place in a sound attenuated chamber. During testing, the child sat comfortably in a recliner chair and watched a movie (DVD or VHS) of his or her choice. The movie soundtrack was presented in free field with the sound level set to < 40 dB SPL, allowing the child to hear the soundtrack via the unoccluded, non-test ear. To enhance compliance, children were accompanied by their parent(s) in the chamber. Children were permitted breaks during testing as needed.

Auditory brainstem responses were collected via the Navigator Pro (Bio-logic Systems Corp., a Natus Company, Mundelein, IL) using BioMAP software (http://www.brainvolts.northwestern.edu/projects/clinicaltechnology.php). All auditory stimuli were presented monaurally into the right ear through insert earphones (ER-3, Etymotic Research, Elk Grove Village, IL, USA). Responses were recorded via three Ag-AgCl electrodes, with contact impedance of ≤ 5 kΩ, located centrally on the scalp (Cz), with the ipsilateral earlobe as reference and forehead as ground. For all data, trials with artifacts exceeding 23.8 μV were rejected online.

Click-evoked Responses

Click-evoked wave V auditory brainstem response latencies were used to assess hearing, as well as for confirmation that the position of the ear insert did not change during the test session. The click stimulus, a 100 μs duration broadband square wave, was presented at a rate of 13/sec. Click-evoked responses were sampled at 24 kHz and were online bandpass filtered from 100-

1500 Hz, 12 dB/octave. Two blocks of 1000 sweeps each were collected at the onset of the session and an additional block of 1000 sweeps was collected at the conclusion to confirm that ear insert placement did not change.

Speech-evoked Responses

Auditory brainstem responses were recorded in response to a 40 ms speech syllable /da/ which was synthesized in Klatt (Klatt, 1980) (Figure 5). In this syllable, the voicing begins at 5 ms and the first 10 ms are bursted. The frequency components are as follows: F_0: 103-125 Hz, F_1: 220-720 Hz, F_2: 1700-1240 Hz, F_3: 2580-2500 Hz, F_4: 3600 (constant), F_5: 4500 (constant); F_2-F_5 comprise what is referred to here as high frequency (HF) information.

The speech-evoked brainstem responses were collected in two different conditions, at conversational speech level in quiet (80 dB SPL) and at the same intensity with simultaneous white background noise (75 dB). The /da/ stimuli were presented with alternating polarity in order to minimize stimulus artifact and cochlear microphonic (Gorga et al., 1985). For each condition, three blocks of 2000 sweeps were collected at a rate of 10.9/sec. A 75 ms recording window (including a 15 ms pre-stimulus period and 20 ms period after stimulus offset) was used. The three response blocks were then averaged together for a final waveform to be used in all analyses. Responses were sampled at 6856 Hz and bandpass filtered on-line from 100-2000 Hz, 12 dB/octave to isolate the frequencies that are most robustly encoded at the level of the brainstem.

Data Analysis

Standardized tests of cognitive (WASI) and language (CELF) abilities were scored according to each test's age appropriate normative range.

The speech-evoked brainstem response was characterized by measures of pre-stimulus neural activity, peak latency and amplitude, the onset response complex' (waves V and A) interpeak duration, slope and amplitude; and FFR amplitude (root-mean-square (RMS), fast Fourier transform amplitude (FFT), stimulus-to-response and quiet response-to-noise inter-response correlations. RMS and FFT measures were performed over the 22-40 ms range, isolating the main peaks of the FFR. Stimulus-to-response correlations were performed in both quiet and background noise conditions by shifting the 13-34 ms range of the stimulus in time until the best correlation with the response was found. This maximum typically occurred at 8.5 ms and aligned the FFR peaks with those of the stimulus waveform. Quiet-to-noise inter-response correlations were performed over the entire response as well as the 11-40 ms FFR range, in each case allowing for up to 2 ms of lag. These measures have been previously described (Russo et al., 2004). A subset of 6 of the original 21 children with ASD were re-evaluated at a second session to establish response reliability over time in children with ASD and demonstrated that the responses reported below were stable over time in these children.

For the statistical analysis, multivariate analyses of variance (MANOVA) tests were conducted, followed by independent two-tailed Student's *t*-tests. Due to limitations inherent in the interpretation of a MANOVA (Tabachnick & Fidell, 2007a), two-tailed independent *t*-tests and

effect sizes were calculated to describe diagnostic group differences (p-values ≤ 0.05 were considered significant and Cohen's d effect sizes > 0.50 were deemed meaningful). A preliminary multivariate analysis of variance ruled out age, sex and mental ability as potential co-factors, with the exception of speech-evoked wave V latency and sex. Reported p-values are those of the t-tests, except for wave V latency. Due to the sex effect, an analysis of co-variance with sex as a co-variate was conducted for wave V latency; therefore, F values are reported. Levene's Test for Equality of Variances was applied to each statistical analysis and, when relevant, the reported p-values reflect corrections based on unequal variances. Pearson's correlations were conducted to establish relationships between neural processing of sound and cognitive/academic measures; the correlation analyses were collapsed across (irrespective of) diagnostic category. Correlations were considered significant if they were moderate to strong ($r \geq 0.35$) with $p \leq 0.05$.

Results

Cognitive and Academic Testing (Table 1)

Although in the normal range, both full scale and verbal mental ability scores from the WASI were significantly lower in the children with ASD ($t(37) = 2.42$, $p = .021$, $d = .82$ and ($t(37) = 3.11$, $p = .004$, $d = 1.01$, respectively), whereas performance mental ability did not differ between groups ($t(37) = 1.09$, $p = .282$, $d = .35$). On the CELF test, children with ASD scored significantly lower on indices of core and receptive language ability ($t(37) = 2.77$, $p = .01$, $d = .87$ and $t(37) = 3.58$, $p = .001$, $d = 1.12$, respectively), but not expressive language ability ($t(37) = 1.80$, $p = .08$, $d = .59$).

Click-evoked Brainstem Response

Wave V latency ranged from 5.15-5.90 ms (M = 5.56 ms, SD = .178), consistent with the previously reported normal range (Gorga et al., 1985; Hood, 1998; Jacobson, 1985). Wave V latency did not differ between groups ($t(37) = 1.46$, $p = .149$, $d = .46$).

Speech-evoked Response Fidelity

The brainstem response to /da/ (Figure 5) consists of 7 transient response peaks (V-O). In the quiet condition, waves were identifiable 100% of the time in the TD group; however 6 children with ASD were missing a wave, specifically in the FFR region, although not always the same wave. Because onset response components are known to become abolished or diminished by background noise even in normal subjects (Russo et al., 2004), these measures were omitted from analyses. Analyses of responses in background noise were restricted to those waves (F and O) which were reliably present in background noise (~90% typical responses). Nevertheless, all sustained response measures (RMS, FFT, and correlations) were evaluated. Despite normal click-evoked responses, the ASD group showed pervasive deficits transcribing speech in quiet and background noise, as was evident in both the onset and FFR portions of the response. Significant group differences are described below. Due to the large number of dependent variables, only the means and standard deviations of measures that were significantly different between groups are reported in Table 9.

Transcription of Phonetic (filter) Aspects of Speech

Wave V and A latencies were significantly delayed in children with ASD (wave V: $F(1,36) = 4.45$, $p = .042$, $d = .87$; wave A: $t(37) = 3.45$, $p = .001$, $d = 1.18$) (Figure 6). Onset response duration was also significantly prolonged in the ASD group ($F(1,36) = 4.67$, $p = .037$, $d = .75$).

Transcription of Fundamental Frequency (source) Aspects of Speech

Quiet response wave D and F latencies were significantly delayed in children with ASD (wave D: $t(37) = 2.47$, $p = .018$, $d = .81$; wave F: $t(37) = 2.62$, $p = .013$, $d = .87$), and wave F amplitude showed a trend toward being reduced ($t(37) = 1.91$, $p = .064$, $d = .61$). Wave F amplitude in background noise was smaller in children with ASD ($t(37) = 2.14$, $p = .039$, $d = .70$). These differences are shown in Figure 7. In the frequency domain, there were no between group differences in F_0 magnitude.

Neural Synchrony in Background Noise: The Sustained Response

Stimulus-to-response-in-background-noise correlations over the FFR (13-34 ms) were lower ($t(37) = 2.41$, $p = .021$, $d = .78$) and the maximum correlation occurred at a shorter lag ($t(37) = 2.03$, $p = .050$, $d = .66$) in the ASD group. Quiet-to-noise inter-response correlations over the entire response (0-49 ms) and restricted to the FFR range (11-40 ms) were also significantly lower in the ASD group ($p < .018$, $d > .80$, both ranges) indicating poorer response fidelity in the ASD group. These findings are consistent with excessive response degradation by background noise in the ASD group relative to the TD controls.

Overall, subcortical transcription of sound was pervasively disrupted in ASD children as was evidenced by both phonetic/filter (delayed onset) and F_0/source deficits (delayed FFR waves D and F and smaller amplitude of wave F).

Relationship Between Neurophysiology and Behavior

Individual measures

Quiet: Performance mental ability was related to wave C amplitude ($r(37) = .38$, $p = .018$) while all scores of mental ability (performance, verbal and full scale) were also related to offset wave O amplitude (performance and verbal: $r(37) = .35$, $p = .028$; full scale: $r(37) = .38$, $p = .018$), such that better scores on tests of mental ability were indicative of larger amplitudes of these transient responses.

Noise: Higher performance and full scale mental ability scores were associated with earlier wave F latency in background noise ($r(37) = .35$, $p = .04$ and $r(37) = .38$, $p = .22$, respectively). Higher core and receptive language indices were associated with greater quiet-to-noise inter-response correlations (11-40 ms range; $r(37) = .36$, $p = .025$ and $r(37) = .35$, $p = .027$, respectively).

Composite Analyses

Based on the variables that differed significantly between groups, we computed four composite scores: onset synchrony in quiet (waves V and A latencies and VA duration), transient responses in quiet (waves V, A, D and F latencies, and VA duration), phase locking in quiet (waves D and

F latencies), and neural synchrony in noise (wave F amplitude, stimulus-to-response-in-background-noise correlations and lag, and quiet-to-noise inter-response correlations). Composite scores were significantly worse in children with ASD compared to TD children on all measures (onset synchrony in quiet: $t(37) = 3.59$, $p = .001$, $d = 1.13$; transient responses in quiet: $t(37) = 3.92$, $p < 0.001$, $d = 1.23$; phase locking in quiet: $t(37) = 3.26$, $p = .003$, $d = 1.03$; neural synchrony in noise: $t(37) = 3.21$, $p = .003$, $d = 1.04$).

Of the composite scores, neural synchrony in noise was the only variable that correlated significantly with behavior. Higher core and receptive language indices were indicative of better resilience in background noise (i.e., greater neural synchrony score) ($r(37) = .53$, $p < .001$ and $r(37) = .36$, $p = .02$, respectively).

Discussion

Summary

To our knowledge, this is the first study to investigate the sensory transcription of both filter and source aspects of speech in ASD, as well as to investigate effects of background noise on brainstem processing in this population. The major finding is that children with ASD demonstrate reduced neural synchrony and phase locking to speech cues in both quiet and background noise at the level of the brainstem, despite normal click-evoked responses. Thus, the sensory transcription of speech is disrupted due to an inability to accurately process both filter cues, which help to distinguish between consonants and vowels, and source cues, which help to determine speaker identity and intent. Similar to recent findings of correlations between

behavioral mental ability and cortical evoked potentials in children (Salmond et al., 2007), our results indicate relationships between speech-evoked brainstem responses and behavior. In quiet, more robust encoding of cues that signify formant transitions and offset (amplitudes of transient waves C and O) were related to better cognitive abilities. In background noise, wave F was correlated with higher cognitive ability. Finally, neural resilience to background noise (quiet-to-noise inter-response correlations; wave F amplitude) was strongly related to better core and receptive language abilities. The transcription of speech in noise remaining consistent with transcription of speech in quiet indicates less of a deleterious effect of background noise and may account for better speech perception in noise. Taken together, these findings support the notion that subcortical biological deficits may underlie social communication problems in ASD.

Click Versus Speech

Although there were differences between the control and ASD groups with respect to brainstem processing of speech stimuli, all subjects exhibited normal processing of non-speech (click) stimuli. This finding is consistent with data from cortical evoked potentials in ASD which indicate that the "speech-ness" of a stimulus predisposes to abnormal processing (Boddaert et al., 2003; Boddaert, Chabane, Gervais et al., 2004; Ceponiene et al., 2003; Jansson-Verkasalo et al., 2003; Kasai et al., 2005; Kuhl et al., 2005) (Lepisto et al., 2007; Lepisto et al., 2005; Lepisto et al., 2006).

The different subcortical encoding of click versus speech stimuli probably reflects the distinctive acoustic and environmental characteristics of the stimuli themselves, as has been the case with

cortical responses (Binder et al., 2000; Liebenthal, Binder, Piorkowski, & Remez, 2003). For example, clicks are brief in duration, have a rapid onset and flat broadband spectral components whereas speech syllables, with their consonant-vowel (CV) combinations, are longer in duration, have ramped onset, and complex, time-varying spectral content. Backward masking (masking of the preceding consonant by an after-coming vowel) is characteristic of CV syllables but not clicks. Whereas the response to a click is limited to the onset, the response to speech syllables includes the FFR which is thought to activate different response mechanisms from onset responses (Akhoun et al., in press; Hoormann et al., 1992; Khaladkar, Kartik, & Vanaja, 2005; Song, Banai, Russo, & Kraus, 2006).

It is also possible that the speech-click response discrepancy reflects different exposure that people have to the two stimuli. Lifelong linguistic and musical experience have been shown to influence brainstem transcription of sounds (Krishnan et al., 2005; Krishnan et al., 2004; Musacchia et al., 2007; Wong et al., 2007; Xu et al., 2006). Thus, speech-like sounds would be expected to invoke experience-related shaping of afferent sensory circuitry unlike clicks, which are laboratory-based and artificial. Taken together, these qualities make speech sounds a better stimulus for investigating language-related auditory processing deficits at the brainstem.

Relationship to Other Language-impaired Populations

Both genetic (Bartlett et al., 2004; Herbert & Kenet, 2007; S. D. Smith, 2007)and behavioral evidence suggest that the language impairment in ASD has features in common with other developmental language and language-based learning disorders, including specific language

impairment (SLI) and dyslexia (Herbert & Kenet, 2007; Oram Cardy et al., 2005; Rapin & Dunn, 2003; S. D. Smith, 2007). Similar behavioral characteristics include both delayed or abnormal language development and problems with pragmatics of language. Physiologically, children with ASD show some similarities in subcortical transcription of sound with a subgroup of children with other language-based learning disorders (e.g., poor readers). Common response traits between the groups are a deficit in onset synchrony and reduced fidelity of the response in background noise with respect to the stimulus (Cunningham et al., 2001; Johnson et al., 2007) (King et al., 2002; Russo, Nicol, Zecker, Hayes, & Kraus, 2005; Wible et al., 2004, 2005). Consistent with known deficits in phonologic awareness, children with other language impairments (such as dyslexia) also demonstrate deficits in offset responses and phase locking to higher frequencies (F_1 range), while fundamental frequency-related (pitch-related) response features are unimpaired. In contrast, children with ASD demonstrate more severe degradation of the response to speech in background noise (reduced quiet-to-noise inter-response correlations), amidst robust encoding of frequencies in the F_1 range and normal offset encoding. Although as a group, children with ASD did not exhibit pitch-related deficits (F_0 amplitude, FFR interpeak latencies), isolation of the subset of children with ASD (n=5) previously identified as having poor brainstem pitch tracking in response to fully voiced speech syllables (Russo et al., in press) reveals pitch processing impairments to the syllable /da/. This subset of children did show prolonged interpeak latencies and reduced amplitude of F_0. Because of the relationship with source cues and prosody, the source-related brainstem deficits in ASD may be associated, to some extent, with behavioral difficulties with prosody. Taken together, these data both provide overlap as well as isolate a different basis for speech transcription interference in brainstem

responses of children with ASD compared to individuals with other language-impairments. The data indicate a more pervasive impairment in speech transcription in ASD – in processes important for extracting timing, pitch, and harmonic information from phonemes. These results invite further investigation of speech-related brainstem processing deficits and the language impairment in ASD.

Implications

These data suggest that subcortical auditory pathway dysfunction may contribute to the social communication impairment in ASD. The mechanisms regulating the relationship between brainstem transcription and language acquisition, perception, and production likely involve corticofugal modulation of afferent sensory function. The malleability of this subcortical transcription process by short-term (Russo et al., 2005; Song, Skoe, Wong, & Kraus, 2007) and life-long (Krishnan et al., 2005; Krishnan et al., 2004; Musacchia et al., 2007; Wong et al., 2007; Xu et al., 2006) experience with sound further supports corticofugal modulation. Additional studies are warranted to further investigate reciprocal (top-down and bottom-up) influences. Consistent with prior work demonstrating deficient neural correlates of pitch in response to a fully voiced speech syllable in the auditory brainstem of children with ASD (Russo et al., in press), this study indicates that the brainstem deficit in ASD affects the transcription of various acoustic cues relevant for speech perception.

An advantage of using speech-evoked brainstem responses in the assessment of auditory function in the ASD population is that responses are objective and can be collected passively and non-

invasively. Moreover, children with abnormal speech-evoked brainstem responses represent the best candidates for auditory training (Hayes et al., 2003; King et al., 2002; Russo et al., 2005). Consequently, brainstem responses may eventually have a place as a diagnostic and outcome measure of the efficacy of auditory training programs in the ASD population.

CHAPTER IV: DEFICIENT BRAINSTEM ENCODING OF PITCH IN CHILDREN WITH AUTISM SPECTRUM DISORDERS

Accepted for Publication in *Clinical Neurophysiology*

N. M. Russo[a,b], E. Skoe[a], B. Trommer[b,c], T. Nicol[a], S. Zecker[a], A. Bradlow[b,d], N. Kraus[a,b,e]

[a]The Roxelyn and Richard Pepper Department of Communication Sciences, Northwestern University, Evanston, IL, [b]Northwestern University Interdepartmental Neuroscience Program, Evanston, IL, [c]Departments of Pediatrics and Neurology, Feinberg School of Medicine, Northwestern University, Chicago, IL and Evanston Northwestern Healthcare, Evanston, IL, [d]Department of Linguistics, Northwestern University, Evanston, IL, [e]Department of Neurobiology and Physiology and Department of Otolaryngology, Northwestern University, Evanston, IL.

Abstract

Objective: Deficient prosody is a hallmark of the pragmatic (socially contextualized) language impairment in Autism Spectrum Disorders (ASD). Prosody communicates emotion and intention and is conveyed through acoustic cues such as pitch contour. Thus, the objective of this study was to examine the subcortical representations of prosodic speech in children with ASD.

Methods: Using passively-evoked brainstem responses to speech syllables with descending and ascending pitch contours, we examined sensory encoding of pitch in children with ASD who had normal intelligence and hearing and were age-matched with typically-developing (TD) control children.

Results: We found that some children on the autism spectrum show deficient pitch tracking (evidenced by increased frequency and slope errors and reduced phase locking) compared with TD children.

Conclusions: This is the first demonstration of subcortical involvement in prosody encoding deficits in this population of children.

Significance: Our findings may have implications for diagnostic and remediation strategies in a subset of children with ASD and open up an avenue for future investigations.

Keywords: auditory brainstem, autism, pitch tracking, prosody

Introduction

Autism Spectrum Disorders (ASD) refers to the cluster of disorders including autism, Asperger Disorder, and Pervasive Developmental Disorder Not Otherwise Specified (PDD-NOS). Impairment in pragmatic (socially contextualized) language is a hallmark of all ASD. Prosodic elements of spoken language, including alterations in pitch, duration and amplitude at the word and phrase levels, convey pragmatic information including the importance of a particular word, the requirement for a response to an utterance, or the speaker's affective state. Whereas aberrant prosodic elements — poor inflection, excessive or misaligned stress, monotonous intonation — are known to characterize the expressive language of individuals with ASD (McCann & Peppe, 2003), less is known about the potential contribution of a neurological source to this receptive prosody deficit. Prosody in autism has been extensively investigated at cognitive and behavioral levels (Hobson, 1986) (Paul, Shriberg et al., 2005; Rapin & Dunn, 2003; Shriberg et al., 2001) but a better understanding of the underlying neurophysiology is warranted. Specifically, subcortical responses to prosodic speech have never been studied in individuals with ASD.

Cortical Processing of Prosody in ASD

Although data addressing brainstem involvement are lacking, studies using cortical evoked potentials in patients on the autism spectrum (specifically Asperger Disorder) have demonstrated deficient encoding of speech and related this deficit to poor receptive prosody. For example, adults with Asperger Disorder who were presented with a woman's name uttered neutrally or with scornful, sad, or commanding affect had relative difficulty identifying the emotional connotations compared with controls, and also showed significant differences in mismatch negativity (MMN, a response reflecting encoding of acoustic change) including longer latencies, smaller amplitudes, and fewer elicited responses(Kujala et al., 2005). In a second study (Korpilahti et al., 2006), boys with Asperger Disorder were presented with a woman's name at two different fundamental frequencies (f_0) to express either tender or commanding affect. Their N1 responses (reflecting stimulus onset) were both delayed and reduced in amplitude compared with controls, and their MMN responses were earlier, larger, and had atypical laterality. The most recent study using the MMN in this population showed an enhanced response (amplitude) in individuals with ASD in a constant-feature discrimination for both pitch and vowel stimuli, whereas this effect disappeared when the condition involved deciphering phonemes with pitch variations (Lepisto et al., 2007). These data are similar to earlier work by Lepisto and colleagues, indicating that adults and children with Asperger Disorder (Lepisto et al., 2006), as well as children with autism (Lepisto et al., 2005), had enhanced MMN responses to sounds that deviated in pitch from the standard stimulus. In this study, both the standard and deviant stimuli had constant pitch

for the duration of the sound. However, they also showed reduced P3a responses (involuntary orienting response) to changes in pitch in speech, albeit non-variant, within the syllable.

Pitch and the Auditory Brainstem

Pitch is the psychophysical correlate of f_0 and is determined by the rate of vocal fold vibration. The auditory brainstem encodes frequency components of speech with high fidelity such that the f_0 and its harmonics can be extracted from the passively-elicited auditory brainstem response (Galbraith, Amaya et al., 2004; Kraus & Nicol, 2005; Krishnan et al., 2005; Krishnan et al., 2004; Musacchia et al., 2007; Wong et al., 2007). Accurate brainstem encoding of the pitch contour of a speech syllable is crucial for producing and perceiving both linguistic meaning (e.g. statement vs. question) and emotional affect in speech.

An emergent body of literature has demonstrated that pitch tracking in the auditory brainstem is experience-dependent, malleable and linked to the processing of higher order cognitive factors such as language and music. For example, adult native speakers of a tonal language (Mandarin) demonstrated more precise brainstem pitch encoding than did non-native speakers (Krishnan et al., 2005; Krishnan et al., 2004). Similarly, brainstem frequency-following responses (FFR) more faithfully encoded stimulus f_0 contour and demonstrated more robust phase locking in musicians than non-musicians (Musacchia et al., 2007; Wong et al., 2007). Finally, brainstem pitch tracking can be improved by short-term training (J. Song et al., 2007).

Because click-evoked auditory brainstem responses have historically been used to detect abnormal auditory encoding of sound in the clinical setting, most existing ASD research assesses the integrity of the auditory brainstem via this method (Klin, 1993; Rapin & Dunn, 2003). However, work from our laboratory has demonstrated that some children with language-based learning problems exhibit deficient brainstem encoding of speech stimuli despite normal encoding of click stimuli (Banai et al., 2005; Johnson et al., 2007; Song et al., 2006). Thus, speech stimuli have been shown to be more sensitive and hence more useful than click stimuli for the detection of subtle abnormalities in the processing of language. This finding could be particularly relevant to children with ASD because the transient and periodic dimensions of speech stimuli convey prosodic as well as phonetic information.

To test the hypothesis that faulty brainstem representation of variations in pitch contributes to the impaired prosody in ASD, we compared responses to speech syllables with descending and ascending pitch contours in a population of children with ASD to those of a control population of typically-developing (TD) children.

Methods

The Institutional Review Board of Northwestern University approved all research and consent and assent were obtained from the parent(s) or legal guardian(s) and the child.

Children were acclimated to the testing circumstances prior to experimental data collection. They were allowed to visit the laboratory and interact with the tester on multiple occasions.

Some children brought an electrode home with them to better familiarize themselves with the neurophysiological procedure.

Participants

Of the 48 children originally recruited for this study, six (all children with ASD) were excluded for the following reasons: abnormal click-evoked brainstem responses (N=2), mental ability below inclusion cutoff (N=1), non-compliance resulting in inability to test (N=1), parental decision to discontinue due to the required time commitment of the study (N=1) and relocation (N=1). Final participants included 21 verbal children with ASD (N=19 boys, 2 girls) and 21 typically-developing children (TD, N=13 boys, 8 girls). Age range was 7-13 years old and mean age (years ± SD) did not differ between groups (9.90 ± 1.921 in ASD vs. 9.95 ± 2.085 in TD; independent two-tailed t-test; t=0.077, p=0.939).

Study participants were recruited from community and internet-based organizations for families of children with ASD. They were required to have a formal diagnosis of ASD made by a child neurologist or psychologist and to be actively monitored by their physicians and school professionals at regular intervals. Parents were asked to supply the names of the examining professionals, their credentials, office location, date of initial evaluation and the specific diagnosis made. These parent-reported diagnoses included autism (n=1), Asperger Disorder (n=7), PDD-NOS (n=1), and a combined diagnosis (i.e., Asperger Disorder/PDD-NOS; n=12). Additionally, parental reports indicated deficient prosody perception in the children with ASD. The diagnosis of ASD was supplemented by observations during testing

such that included subjects were noted to have some or all of the following: reduced eye contact, lack of social or emotional reciprocity; perseverative behavior; restricted range of interests in spontaneous and directed conversation during testing set-up; repetitive use of language or idiosyncratic language; abnormal pitch, volume, and intonation; echolalia or scripted speech; and stereotyped body and hand movements. Diagnosis was also supplemented by an internal questionnaire that provided developmental history, a description of current symptoms, and functional level at time of entry into the study.

Further inclusion criteria for both TD and ASD groups were 1) the absence of a confounding neurological diagnosis (e.g. active seizure disorder, cerebral palsy), 2) normal peripheral hearing as measured by air threshold pure-tone audiogram and click-evoked auditory brainstem responses and 3) a full-scale mental ability score whose confidence interval included a value ≥80.

Hearing screening

Normal hearing thresholds and click-evoked wave V latencies confirmed normal hearing status and were required for inclusion in this study. On the first day of testing, children underwent a screening for normal bilateral peripheral hearing (≤20 dB HL) for octaves between 250 and 8000 Hz via an air conduction threshold audiogram on a Grason Stadler model GSI 61. Children wore insert earphones in each ear and were instructed to press a response button every time they heard a beep. At each subsequent test session, children were

required to pass a follow-up hearing screening at 20 dB HL for octaves between 125 and 4000 Hz. These screenings were conducted using a Beltone audiometer and headphones.

Mental and language ability assessment

The Wechsler Abbreviated Scale of Intelligence (WASI, four subtests) (Woerner & Overstreet, 1999) was administered to screen for intellectual ability and provided scores of verbal, performance, and full-scale mental ability (Figure 8; mean and standard error values are plotted). A full-scale mental ability score whose confidence interval included a value ≥80 was necessary for inclusion in the study. Performance and verbal mental ability scores were recorded, but not used as inclusion criteria. Additionally, the Clinical Evaluation of Language Fundamentals-4 (CELF) (Semel et al., 2003) was administered to provide indices of core, expressive and receptive language abilities (Fig. 8). Performance on the CELF was not used as a study inclusion criterion.

Neurophysiology Recording and Stimuli

All neurophysiological recordings took place in a sound attenuated chamber. During testing, children sat comfortably in a recliner chair and watched a video of their choice while experimental stimuli were delivered monaurally to their right ear. The movie soundtrack was presented in free field with the sound level set to <40 dB SPL, allowing the child to hear the soundtrack via the unoccluded, non-test ear. Children were instructed to ignore the sounds being delivered to their right ear and attend to the movie. Because brainstem responses were collected passively, the results were not biased by attention and cognitive abilities, an

important consideration with an ASD population. To ensure compliance of the child and to alert the tester of any complications during testing, the child's parent(s) sat with the child in the chamber. At any time, if the child chose to discontinue testing or take a break during testing, s/he was allowed to do so without penalty.

Auditory evoked responses were recorded via three Ag-AgCl scalp electrodes located centrally (Cz), with an earlobe reference and forehead as ground; all electrodes maintained a contact impedance of ≤5 kΩ. Stimuli were presented via ER-3 insert earphones (Etymotic Research, Elk Grove Village, IL, USA).

The click stimuli (100 µs duration square waves with broad spectral content) were presented at 80.3 dB SPL at a rate of 13/sec Hz. Click-evoked responses (10.66 ms recording window) were digitally sampled at 24 kHz and were online bandpass filtered from 100-1500 Hz, 12 dB/octave. Trials with artifacts exceeding 23.8 µV were rejected online. Two blocks of 1000 non-rejected sweeps each were collected at the beginning of the neurophysiologic test session and an additional block of 1000 sweeps was collected at the conclusion to confirm that ear insert placement did not change during testing. A criterion of being within one cursor-click on the recording window (~0.04 ms) was implemented. In most cases there was no change and in no subject was there more than the allowable difference between initial and final click replications.

The speech syllables were created from a natural spoken /ya/ syllable (fully voiced, flat pitch contour) that was produced by a native English speaking female and subsequently manipulated in Praat (Boersma & Weenink, 2004). The speech sample was duration normalized to 230 ms before digitally manipulating the fundamental frequency (f_0) contour of the original production to create the descending and ascending reciprocal pitch contours (descending: 220-130 Hz; ascending: 130-220 Hz). Descending and ascending contours were chosen to provide a basic model of statement versus question. Because the stimuli originated from the same speech token, all acoustic parameters, with the exception of f_0, were identical.

Speech stimuli were presented at 60 dB SPL in alternating polarities. Alternating polarities were presented in order to minimize stimulus artifact and cochlear microphonics (Gorga et al., 1985). To avoid any potential confound of an anticipatory response, the stimuli were presented in random order with a variable interstimulus interval of 51 ms (±16 ms) (Neuroscan, Stim, Compumedics, El Paso, TX). Speech-evoked responses were recorded (Neuroscan, Scan, Compumedics) at sampling rate of 20,000 Hz. Two replications of 1200 sweeps/polarity (total, 4800) were recorded for each syllable. Trials with artifacts greater than 35 μV were rejected offline. On average, 92% of the trials (~8800/9600 sweeps; range: 6207-9567 sweeps) remained after artifact rejection.

Analyses

Click-evoked brainstem responses

Wave V latency was identified for each subject and needed to fall within the normal range for 80 dB SPL clicks. Delayed wave V latency was used as an exclusionary criterion because latencies beyond the normal range may indicate other confounding deficits.

Speech-evoked brainstem responses: Pitch tracking in the auditory brainstem

Speech-evoked response waveforms were averaged offline in Neuroscan with a recording time window spanning from 50 ms prior to the onset of the stimulus until 20 ms past the offset. Responses were bandpass filtered offline from 80-1000 Hz with a 12 dB/octave rolloff to isolate the frequencies that are most robustly encoded at the level of the brainstem. For the purpose of calculating signal-to-noise ratios (SNR), a single waveform per subject representing the non-stimulus evoked activity was created by averaging the neural activity prior to stimulus onset.

For all analyses, measures were first assessed in stimulus-specific responses and then averaged across stimuli to obtain a single number for each measure for each participant. This combination was possible because the same patterns were observed with both descending and ascending /ya/ conditions. Thus, the combined-stimulus averages are reported here. All pitch-tracking analyses were performed using routines coded in Matlab 7.0.4 (The Mathworks, Natick, MA).

Frequency-following response pitch contours were extracted for each subject for the f_0 and second harmonic (H2) and analyzed with respect to the frequency contours of the stimuli.

Pitch tracking and phase locking were described by measures of Frequency Error, Slope Error, and Pitch Strength. Frequency Error represented the accuracy of pitch encoding over the duration of the stimulus. Slope Error measured the degree to which the shape of the pitch contour was preserved in the response. Pitch Strength, a measure of response periodicity, indicated the robustness of neural phase locking to the stimulus f_0 contour. Due to limitations of the autocorrelation method used for calculating Pitch Strength, H2 was assessed only by Frequency and Slope Error.

Pitch-tracking measures were derived using a sliding window analysis procedure. A 40-ms window was slid across the FFR in 1 ms increments, and an FFT and autocorrelation were computed on each 40-ms portion of the FFR. The window was shifted 190 times and this produced a total of 191 spectral and autocorrelational estimates. The time period encapsulated by each shift of the 40 ms window is referred to as a time bin. In the pitch-tracking and pitch strength plots, the time indicated on the *x*-axis refers to the midpoint of each 40-ms time bin. A short-term Fourier transform was calculated for each Hanning-windowed bin. The resulting spectrogram gave an estimate of spectral energy over time. The f_0 and H2 contours were extracted from the spectrogram by finding the frequency (between 0 and 300 Hz for f_0 and 260-440 Hz for H2) with the largest spectral magnitude for each bin. Spectral peaks that did not fall above the noise-floor (SNR<1) were excluded as possible f_0 or harmonic candidates. The same short-term spectral analysis procedure was applied to the stimulus waveforms (methods for f_0 extraction follow (Musacchia et al., 2007; Wong et al., 2007)).

The three measures of pitch tracking were calculated as follows: To obtain a measure of Frequency Error, the absolute Euclidian distance between the stimulus and response f_0 and H2 (respectively) at each time bin was calculated and then averaged across all 191 bins. Slope Error represented the absolute difference between the slopes of the stimulus and response pitch tracking regression lines. For this measure, the extracted f_0 and H2 data points were fit to a linear model from which a regression line was calculated. The slope of the regression line was recorded and compared to the slope of the regression line created from the stimulus waveforms (f_0: descending stimulus, m=–440 Hz/sec; ascending stimulus, m=460 Hz/sec; H2: descending stimulus, m=–880 Hz/sec; ascending stimulus, m=920 Hz/sec). For calculating all pitch-tracking variables, stimulus measurements were derived from a recording of the original stimuli as presented through the Neuroscan and Etymotic equipment, as this recorded output waveform is an accurate representation of what the participants actually heard. Subtle differences between input and output stimulus waveforms account for the slight deviation in above-reported slopes of the descending and ascending stimuli. (See Fig. 9)

The third measure of pitch tracking, Pitch Strength, was derived using a short-term autocorrelation method. This method is used to determine signal periodicity over time wherein a signal is compared to a time-shifted copy of itself. The time-shift is quantified in terms of lag (ms). For each time lag, a correlation r-value, representing the degree of signal periodicity or pitch strength, is calculated (expressed as a value between -1 and 1).

Fundamental frequency is calculated from the autocorrelation function (r-value vs. lag) by finding the fundamental period — the time lag needed to obtain the correlational maxima — and taking the inverse (frequency = 1/period; e.g., 1/15 ms =66.67 Hz). Because there was no interest in frequencies below 67 Hz, the lag was limited to15 ms.

For the stimulus, the fundamental period of each time bin was recorded. The Pitch Strength of each response bin was quantified as the r-value corresponding to the fundamental period of the stimulus at the corresponding time bin; larger r-vales indicated more periodic time frames. Similar to Frequency Error and Slope Error, Pitch Strength was the average r-value across the 191 bins. The reported mean r-values were converted to Fisher z'-scores for all statistical analyses. Running autocorrelograms (Fig. 10) (Krishnan et al., 2005; Krishnan et al., 2004; Musacchia et al., 2007; Wong et al., 2007) were generated as a means of visualizing and quantifying periodicity and Pitch Strength variation over the course of the response. The x- and y- axes are time and lag, and the third dimension, Pitch Strength, is plotted using a color continuum from black to white, with brighter colors representing higher correlations.

Composite score

To comprehensively quantify the deficit in pitch tracking, Frequency Error of f_0, Pitch Strength and Frequency Error of H2 scores were transformed into z-scores and then averaged together to obtain a composite pitch-tracking score for each subject. To account for the fact that lower values were better for Frequency Error, while higher values were better for Pitch

Strength, Pitch Strength z-scores were first multiplied by a factor of negative one before being entered into the composite score calculation.

Statistical Analyses

A one-way analysis of variance (ANOVA) was used to evaluate group differences in click-evoked response latencies; the two-tailed result is reported because no differences were expected since all children met our inclusion criterion. Multivariate analyses of variance (MANOVA) were conducted between groups to test the hypothesis that sensory encoding of acoustic cues of prosody in speech (here, pitch and harmonic contour) is disrupted in children with ASD. Dependent variables included Frequency Error, Slope Error, and Pitch Strength; diagnosis was the fixed factor. Due to limitations inherent in the interpretation of a MANOVA (Tabachnick & Fidell, 2007a), one-tailed independent t-tests (because our pitch-tracking results were hypothesis-driven) and Cohen's *d* effect sizes were calculated to describe diagnostic group differences (p-values ≤ 0.05 and $d \geq 0.50$ were required to be considered significant). Levene's Test for Equality of Variances was applied to each statistical analysis and, when relevant, the reported p-values reflect corrections based on unequal variances. The non-parametric Kruskal-Wallis statistic was used for subgroup comparisons due to the smaller number of subjects in these groups.

Results

Age, sex, and intelligence considerations

Because of the variability in age and intelligence, we considered these variables in preliminary statistical analyses. Further, due to the greater incidence of ASD in males versus females, our ASD group included a majority of male participants. Since sex differences can occur in brainstem responses (Jerger & Hall, 1980; Rupa & Dayal, 1993), we also evaluated effects of sex. The distribution of age did not vary between groups and therefore it is unlikely to be a contributing factor to any of the differences we report (χ^2=3.652, p=0.724). There were no significant relationships between age, sex or mental ability with any of the neurophysiological measures (Pearson's r-value≤0.093 p≥0.557, all tests). Finally, although there were no significant correlations, preliminary MANOVA ruled out age, sex and mental ability as co-variates for the neurophysiological measures because they were not statistically significant. Thus, subsequent analyses were conducted without these co-variates.

Brainstem responses to click stimuli

All children exhibited normal brainstem responses to click stimuli; there were no between group differences (ASD mean latency=5.6 ms (SD=0.19), TD mean=5.6 ms (SD=0.17); ANOVA, $F_{(1,40)}$=0.772, p=0.385). As a combined group, the TD and ASD wave V latencies ranged from 5.15-5.90 ms, with TD responses ranging from 5.28-5.90 ms. These results were consistent with their normal pure tone audiometric hearing thresholds (≤20 dB HL) and indicated normal encoding of the onset of transient acoustic stimuli.

Encoding f_0

Despite demonstrating normal brainstem responses to click stimuli, children with ASD demonstrated deficient encoding of pitch in speech compared with TD children. Frequency Error was compared between groups and the ASD responses demonstrated less accurate pitch tracking (TD mean (SD)=8.52(2.201) Hz; ASD=10.10(2.912); t=1.99, p=0.027; *d* =0.61; Figs. 2, 4). Slope Error indicated a trend toward greater error in the ASD group (TD=30(20) Hz/sec; ASD=50(44); t=1.58, p=0.063; *d*=0.59; Figs. 9, 11). Further, Pitch Strength autocorrelations were significantly higher in TD responses (TD mean (SD) r=0.39(0.198); ASD mean (SD) r=0.30(0.159); t=1.96, p=0.0465; *d*=0.56; Figs. 10, 11). Lower Frequency Error and higher Pitch Strength values indicated that the stimulus f_0 contour was better preserved and more robustly encoded in the brainstem responses of TD subjects. (See Table 10)

Encoding harmonics

Frequency Error and Slope Error of H2 were also compared; because an autocorrelation is not meaningful for the harmonics, Pitch Strength was not calculated. ASD responses demonstrated greater Frequency Error (TD=13.43(2.071) Hz; ASD=15.06(2.392); t=2.368, p=0.02; *d*=0.73), but Slope Error did not differ between groups.

Composite score and subgrouping of participants

The composite score, described above, was calculated for each participant to provide a comprehensive measure of pitch encoding deficits in the brainstem. TD responses demonstrated significantly better encoding of the pitch contour than ASD responses (TD

z=0.00(0.790); ASD z=0.68(0.888), t=2.636, p=0.012; *d*=1.15; Fig. 11). Using this composite score, we were able to isolate 5 children with ASD (~20%) who demonstrated pitch-encoding deficits greater than 1.65 standard deviations (accounting for 95% of the variance). Therefore, children with ASD were classified as deficient pitch trackers ("ASD OUT," n=5) or typical pitch trackers ("ASD IN," n=16) on the basis of their composite scores. The ASD OUT group included three children with Asperger Disorder, one with PDD-NOS, and one with "Autism Spectrum Disorder plus Sensory Integration Disorder".

Encoding f_0

The individual pitch-tracking measures were re-assessed (Table 10) and revealed that the overall diagnostic group difference reported previously was driven by the ASD OUT group whereas the ASD IN group demonstrated averages comparable to the TD group (Fig. 11). There was a significant group difference in Frequency Error (Kruskal-Wallis test, H(2)=10.415, p=0.005) and Pitch Strength (H(2)=7.337, p=0.026), Slope Error did not reach significance using this categorization (H(2)=2.608, p=0.271). Follow-up one-tailed Mann-Whitney tests showed that the TD and ASD IN groups did not vary significantly on any measure, whereas the ASD OUT group differed significantly from both TD and ASD IN groups on both Frequency Error (U=6.0, p=0.001and U=5.0, p=0.002, respectively) and Pitch Strength (U=13.0, p=0.008 and U=12.0, p=0.019, respectively).

Encoding harmonics

Kruskal-Wallis tests indicated a significant group difference in Frequency Error for encoding of H2, but not in Slope Error of H2 (H(2)=11.472, p=0.003 and H(2)=0.397, p=0.820,

respectively). Follow-up one-tailed Mann-Whitney tests showed that the TD and ASD IN groups did not differ in harmonics encoding, while the ASD OUT group had lower Frequency Error of H2 compared to both the TD and ASD IN groups ($U=5.0$, $p=0.001$ and $U=4.0$, $p=0.001$, respectively).

Relationship to Behavior

ASD subjects had significantly lower scores than TD subjects on both mental ability and language testing ($p<0.025$, all tests), with the exception of performance mental ability ($p=0.133$), for which the ASD group scored similarly to the TD group (Fig. 8). Mann-Whitney tests between the ASD IN and ASD OUT group revealed no significant differences on the behavioral tests ($U=22.5$-32.50, $p\geq0.153$, all tests). There were no significant relationships (Pearson's $r\leq0.421$, $p\geq0.061$, all tests) between pitch tracking in the brainstem and measures of language and mental ability in either diagnostic group or the ASD IN subgroup; it was not possible to evaluate meaningful correlations in the ASD OUT group due to the small group size.

Pitch Tracking Test-rest Reliability

As children with ASD represent a difficult to test population, we were interested in the stability of these results across multiple test sessions. In a separate study (unpublished data), measures of pitch tracking were evaluated for stability from test to follow-up re-test session in six of the original 21 children with ASD (all ASD IN children). With only six children, we chose to conduct a non-parametric paired test (Wilcoxon Signed Ranks test) to be more

conservative. Data indicated no significant changes in f_0 Frequency Error (Z=–0.314, p=0.753), Slope Error (Z=–0.105, p=0.917) and Pitch Strength (Z=–0.105, p=0.917) or in H2 Frequency Error (Z=–0.105, p=0.917) or Slope Error (Z=–0.677, p=0.498) which indicate the reliability and stability of this response.

Discussion

Using speech syllables with variable pitch, we have demonstrated deficient brainstem encoding of pitch in a subgroup of verbal children with ASD. Specifically we found that these children with ASD had aberrant, non-direction-specific pitch tracking (increased frequency and slope error) and reduced neural phase locking to the stimulus (poorer autocorrelations) compared to TD children. These results were detected in children over a restricted age range, with normal peripheral hearing and brainstem conductions times, full scale intelligence scores >80 and without confounding neurological impairment. Because the diagnoses of children in both the ASD IN and ASD OUT groups varied, diagnosis alone was not a distinguishing factor of good or poor pitch tracking. Nevertheless possible effects of diagnoses should be investigated more thoroughly in future work. That only a subset of our population showed abnormalities in the auditory brainstem is consistent with the findings of other investigators (Maziade et al., 2000; Rosenhall et al., 2003) and also consistent with the known heterogeneity within and between diagnostic categories of the autism spectrum (Freitag, 2007). Both the ASD IN and the ASD OUT groups met criteria for ASD, and thus would not be predicted to differ on the behavioral measures that were tested. Neither the WASI nor the CELF specifically target deficits in prosody perception. That they did not

differ in language testing but did differ in FFR is, in our view, a reflection of the greater sensitivity of the electrophysiologic testing. Because the brainstem paradigm is passive, quantifiably poor pitch tracking in the FFR may be more conspicuous than in behavioral tests, during which participants may use other cues and tools to compensate for this deficit. Thus, it is possible that better designed behavioral tests of receptive and expressive prosody may correlate with the deficits in the FFR**.**

Within speech signals, the f_0 and its harmonics are important for conveying affect (Patel, Peretz, Tramo, & Labreque, 1998; Schon, Magne, & Besson, 2004). In a typical system, the auditory brainstem robustly extracts and encodes the pitch contour from the speech signal. In brainstem responses of children with ASD, frequency encoding was non-specific, non-periodic and diffuse such that the most robustly encoded frequency did not correspond to the pitch contour of the stimulus. Thus, in many cases, the f_0 contour was not registered by the brainstem. This raises the possibility that poor brainstem representation of f_0 contour may underlie poor recognition of f_0 as a significant acoustic cue. Although some caution is advised due to our small study sample, our data are consistent with the idea that receptive prosody deficits, and by inference, possibly also expressive prosody deficits, stem from an inability to passively encode and transmit variable pitch contours beginning in the auditory brainstem in some patients.

Brainstem Deficits and Cortical Connections in ASD

Clinical neurophysiology

Several prior studies have examined the integrity of the auditory brainstem in children with ASD and some have reported aberrant brainstem responses to non-speech stimuli (reviewed in (Klin, 1993; Rapin & Dunn, 2003). McClelland and colleagues (1992) found prolonged brainstem transmission times in response to pulse stimulation in mentally-handicapped individuals with ASD (ages 3-23 years) and attributed the delay to maturational defects in myelination. Maziade and colleagues (2000) reported increased inter-peak latencies between waves I-III and I-V using click-evoked brainstem responses in 73 children with ASD (ages 2-12 years) who were compliant for the study and had otherwise normal hearing. The authors concluded that the slowed conduction time could be attributed to reduced myelination, although they also postulated cerebellar degeneration, hyperserotonemia – or a combination of these abnormalities at the brainstem. Similarly, Rosenhall and colleagues (2003) found increased click-evoked brainstem conduction times in just over half of the 153 tested individuals with ASD (ages 4-20 years) although in this study, about 8% of their subjects had hearing loss. That study included some children with mild or severe mental retardation and it was not reported how many of those cases had abnormal brainstem responses. In contrast to these studies, Tharpe and colleagues (2006) did not find sensory encoding deficits at the level of the brainstem in a study of 22 children with ASD (ages 3-10 years). Although click-evoked brainstem responses were normal, pure tone thresholds were atypical in half of their subjects, suggesting that these children might represent a unique subgroup of children with ASD.

These prior brainstem studies employed a relatively restricted stimulus repertoire (i.e., only clicks or pulses), which only allow for investigation of latency and amplitude variations. Our study evaluated frequency encoding in speech in subjects who demonstrated normal brainstem responses to clicks. As in the present investigation, most of the studies report that only subsets of their children show deficiencies. Thus, any discrepancy between studies could be due either to different mechanisms of auditory pathway dysfunction in various subsets of children with ASD or the different mechanisms of processing clicks versus speech (Hoormann et al., 1992).

The Neuro-anatomic Basis

Brainstem development

Experience-dependent postnatal pruning occurs in multiple subcortical components of the normal auditory system (e.g., lateral superior olivary nucleus, lateral lemniscus, and inferior colliculus)(Gabriele, Brunso-Bechtold, & Henkel, 2000; Henkel, Gabriele, & McHaffie, 2005; Sanes & Constantine-Paton, 1985; Sanes & Friaf, 2000) such that irregularities in this process may underlie disordered connectivity within the brainstem and between the cortex and brainstem. For example, in the lateral superior olivary nucleus, the postnatal depolarization of inhibitory input allows for elaboration of pre- and post-synaptic connections whereas hyperpolarization leads to elimination of connections and the balance thus promotes refinement of auditory pathways (Sanes & Friaf, 2000). Additionally, abnormal early auditory input affects post-natal pruning in the lateral lemniscus and inferior

colliculus which is necessary for spectral and temporal auditory function and frequency tuning (Henkel et al., 2005; Sanes & Constantine-Paton, 1985).

Prior clinical and animal research models have implicated deficits in brainstem maturation and development in ASD. Data from magnetic resonance imaging in individuals with ASD (Hashimoto et al., 1993; Hashimoto et al., 1995), and experiments exploiting genetic defects in an animal model (Rodier, 2000; Rodier, Ingram, Tisdale, & Croog, 1997), point to atypical embryological development (deficient maturation) and a smaller brainstem. Hashimoto et al. (1995) and McClelland et al. (1992) also suggested maturational myelin-related deficits at the brainstem that may affect either projections to the limbic system or the auditory cortex (reduced long-range connectivity to the cortex), with fewer ascending projections. Together, these studies provide evidence that the brainstem is implicated in ASD and that the brainstem frequency-following response may be used as a marker for one neuropsychological deficit.

Neuro-anatomic deficits in brainstem-cortical connections in ASD

Disrupted connections between the brainstem and cortex, as well as deficient sensory encoding of speech within cortex (Ceponiene et al., 2003) (Boddaert, Chabane, Gervais et al., 2004; Flagg, Cardy, Roberts, & Roberts, 2005), may account for the auditory processing impairment in individuals with ASD. Anatomical differences in cortical microarchitecture, including decreased long-range connectivity coupled with greater local neuronal proliferation (increased numbers and densely packed neurons), have been linked to autism (Baron-Cohen et al., 2005; Courchesne & Pierce, 2005; Wickelgren, 2005). Because auditory connections

are reciprocal, impaired encoding of pitch contour at the brainstem may affect cortical encoding in a feed-forward fashion by propagating to the ascending auditory pathway (Galbraith, Gutterson et al., 2004). Conversely, because cortical modulation helps shape brainstem encoding and enhances signal processing (Boylan, Blue, & Hohmann, 2007; Galbraith, Olfman, & Huffman, 2003; Suga et al., 2000; Yan & Suga, 1996; Yu, Sanes, Aristizabal, Wadghiri, & Turnbull, 2007), it is plausible that faulty brainstem representation of sound may arise, at least in part, from the lack of optimal top-down, corticofugal engagement of auditory pathway activity. Supporting the theory of disrupted corticofugal function in ASD, Boylan and colleagues (2007) discuss converging evidence (using immunochemistry and autoradiography) implicating abnormal cortical innervation, atypical (or absent) pruning and reorganized sensory maps resulting in perceptual processing deficits in their rodent model of autism. In both "bottom up" and "top down" scenarios, inaccurate input from the brainstem could ultimately contribute to defective cortical encoding of speech prosody in the auditory cortex, and limit comprehension of linguistic affect.

Implications

Brainstem malleability

Brainstem function for speech and music has been shown to be malleable with short term training (Russo et al., 2005; J. Song et al., 2007) and sharpened by lifelong auditory experience with language (Krishnan et al., 2005; Xu et al., 2006), and music (Musacchia et al., 2007; Wong et al., 2007) likely through corticofugal mechanisms. For example, Krishnan and colleagues found that Mandarin speakers had more finely tuned pitch encoding

in the brainstem, indicating that brainstem pitch tracking is modulated by language experience (Krishnan et al., 2005; Xu et al., 2006) and musicians have been shown to exhibit enhanced brainstem encoding of both speech and music (Musacchia et al., 2007). Further, although they do not show the same deficits with expressive and receptive prosody, some children with language-based learning problems have brainstem deficiencies encoding acoustic aspects of speech (Banai et al., 2005; Johnson et al., 2007; Wible et al., 2004). Following auditory training, components of the brainstem FFR, of which f_0 encoding is a major part, become less "noisy" (fewer non-stimulus related spectral peaks) after auditory training (Russo et al., 2005), a finding that may have direct application to children with ASD.

Because prosody is often considered the "music of language," music therapy may facilitate pitch learning in language (Schon et al., 2004). Kellerman and colleagues (Kellerman, Fin, & Gorman, 2005) suggest that the repetitive nature of music is attractive to individuals with ASD and it has also been proposed that the technical aspects of music appeal to individuals with ASD (Levitin, 2006). Some benefits of music therapy have been reported in treating the communication deficit in ASD; case studies have shown that music therapy improved both production and interpretation of others' intonation (Hoelzley, 1993; S. B. Miller & Toca, 1979). In addition, enhanced brainstem encoding of pitch with long-term musical training has been shown for both speech and music (Musacchia et al., 2007; Wong et al., 2007). Extended exposure to music appears to sharpen the auditory encoding of speech containing prosodic pitch contours. The malleability of brainstem encoding and its enhancement with

musical training support the view that auditory training aimed at improving pitch tracking, including music training, may provide therapeutic intervention for some children with ASD.

Summary

The brainstem response to speech is a passively-elicited, non-invasive objective index of brainstem encoding of key linguistic cues. Using this response, we have shown that some children with ASD demonstrate marked deficiencies in pitch tracking, offering an attractive candidate mechanism for their deficient receptive prosody. Because the brainstem response matures early, this paradigm could conceivably be utilized to screen for severe deficits in pragmatic language in infants or young children, which may be indicative of early symptoms of ASD.

Several modifications can be anticipated to improve the precision of our approach to the study of the neurophysiology of language impairment in autism. These include the expansion of our study paradigm to include aspects of prosody encoding other than pitch (variations in stress/emphasis), aspects of speech encoding other than prosody (e.g., consonant-vowel syllables with invariant pitch,), standardized behavioral measures of receptive prosody impairment and, finally, more precise tools for clinical classification of subjects (the Autism Diagnostic Observation Schedule (Lord et al., 2000; Lord et al., 1989) and Autism Diagnostic Interview-Revised (Le Couteur et al., 1989; Lord, Rutter, & Le Couteur, 1994)). Together these modifications are likely to improve our ability to characterize language deficits in children with ASD and further work that incorporates this

paradigm may also produce a viable neurophysiologic marker for subtyping these children in conjunction with genetic and behavioral analyses.

Acknowledgements

We would like to thank the children who participated in this study and their families. We would also like to Erika Skoe for her contribution to this research, particularly in software development. Financial Interests: This research was supported by NIH R01 DC01510. The authors declare that they have no competing financial interests.

CHAPTER V: AUDIO-VOCAL SYSTEM REGULATION IN CHILDREN WITH AUTISM SPECTRUM DISORDERS

Accepted for Publication in *Experimental Brain Research*

Nicole Russo[a,b]; Charles Larson[a,b,d]; Nina Kraus[a-d]

[a]The Roxelyn and Richard Pepper Department of Communication Sciences, [b]Northwestern University Interdepartmental Neuroscience Program, [c]Department of Neurobiology and Physiology, [d]Department of Otolaryngology, Northwestern University, Evanston, Illinois.

Abstract

Do children with autism spectrum disorders (ASD) respond similarly to perturbations in auditory feedback as typically-developing (TD) children? Presentation of pitch-shifted voice auditory feedback to vocalizing participants reveals a close coupling between the processing of auditory feedback and vocal motor control. This paradigm was used to test the hypothesis that abnormalities in the audio-vocal system would negatively impact ASD compensatory responses to perturbed auditory feedback. Voice fundamental frequency (F_0) was measured while children produced an /a/ sound into a microphone. The voice signal was fed back to the subjects in real time through headphones. During production, the feedback was pitch shifted (–100 cents, 200 ms) at random intervals for 80 trials. Averaged voice F_0 responses to pitch-shifted stimuli were calculated and correlated with both mental and language abilities as tested via standardized tests. A subset of children with ASD produced larger responses to perturbed auditory feedback than TD children, while the other children with ASD produced significantly lower response magnitudes. Furthermore, robust relationships

between language ability, response magnitude and time of peak magnitude were identified. Because auditory feedback helps to stabilize voice F_0 (a major acoustic cue of prosody) and individuals with ASD have problems with prosody, this study identified potential mechanisms of dysfunction in the audio-vocal system for voice pitch regulation in some children with ASD. Objectively quantifying this deficit may inform both the assessment of a subgroup of ASD children with prosody deficits, as well as remediation strategies that incorporate pitch training.

Keywords: autism, vocal production, auditory feedback

Introduction

Autism spectrum disorders (ASD) are developmental disorders in which one of the primary indicators is language impairment with respect to social communication, including expressive control of prosody in speech. Variations in prosody distinguish declaratory statements from interrogatories, give clues to the speaker's emotional tone of voice, and indicate when words or statements begin and end. Many individuals with ASD have problems with prosody in speech, including the perception of pitch and production (regulation) of changes in voice fundamental frequency (F_0) over time(McCann & Peppe, 2003; Rapin & Dunn, 2003). As a behaviorally diagnosed spectrum disorder, the ASD population remains densely heterogeneous (Freitag, 2007). Thus, in the current absence of objective measures for diagnosis, there is a need to identify viable biological and physiological diagnostic markers (Filipek et al., 2000) This task can be accomplished by

investigating each core symptom of ASD separately. The focus of this study is the regulation of voice F_0 and its relationship to language impairment in ASD.

Language development is significantly disrupted in ASD. Some children with ASD are non-verbal; others develop language, but then experience a loss (or regression) of language. Finally, still other children develop language later than expected. The speech of verbal children with ASD is often monotonous, echolalic or stereotypic, inappropriately stressed, or emotionless (Boucher, 2003; Rapin & Dunn, 2003; Shriberg et al., 2001; Siegal & Blades, 2003). Appropriate voice F_0 modulation is crucial for successful social interaction as it imparts information about the subject's state of mind, emotion, or intent. Thus, due to the abnormal prosody of speech in children with ASD, conversation with peers is often strained (McCann et al., 2007; Paul, Shriberg et al., 2005).

Prior studies have investigated the potential relationship between the language impairment in ASD and the auditory processing of sound and have shown some evidence for peripheral, subcortical, and cortical abnormalities. Evaluation of evoked otoacoustic emissions in children with autism revealed atypical asymmetry in the medial olivocochlear system, as well as a decrease in otoacoustic emissions with age (within children and adolescents), which was not seen in the control children (Khalfa et al., 2001). In contrast, Gravel and colleagues (Gravel et al., 2006) showed no behavioral differences in the peripheral auditory system in high-functioning children with autism. Tharpe and colleagues evaluated both peripheral audiometry and brainstem function in children with autism (Tharpe et al., 2006). Pure tone

thresholds were atypical in half of their subjects, yet this difference was not corroborated by click- or tone-evoked auditory brainstem response recordings. Although Tharpe and colleagues (2006) did not find brainstem deficits, other studies of brainstem integrity have identified aberrant function (Klin, 1993; Maziade et al., 2000; McClelland et al., 1992) (Rapin & Dunn, 2003; Rosenhall et al., 2003; Russo et al., in press). In one study investigating brainstem transcription of F_0 contour in speech in children with ASD, deficient pitch tracking was identified in only a subset of those children, while brainstem function was normal in the other children with ASD (Russo et al., in press). Further, there is ample evidence for deficient or atypical cortical processing of speech or speech-like stimuli associated with ASD (Boddaert et al., 2003; Boddaert, Chabane, Belin et al., 2004; Ceponiene et al., 2003; Gervais et al., 2004; Jansson-Verkasalo et al., 2003; Kasai et al., 2005; Lepisto et al., 2005; Lepisto et al., 2006; Wang, Dapretto, Hariri, Sigman, & Bookheimer, 2001), including reports of deficient cortical processing specific to prosody (Erwin et al., 1991; Korpilahti et al., 2006; Kujala et al., 2005; Wang et al., 2001). Even amidst these recent findings, much of the physiology behind the language impairment and characteristic speech production patterns in ASD is still unmapped.

Adequate hearing is critical for speech development. Although little is known about the role of auditory feedback in speech production in individuals with ASD, ample evidence from studies of individuals with post-lingual deafness and cochlear implants (CI) indicate the necessity of auditory feedback for vocal control of loudness and pitch (Campisi et al., 2005; Hamzavi et al., 2000; Higgins et al., 1999; H. Lane et al., 1997; Leder et al., 1987; Monini et

al., 1997; Perkell et al., 1992; Svirsky et al., 1992). People who are pre-lingually deafened almost never develop clear speech. Those who are post-lingually deafened show marked deterioration in control of prosodic features of speech (such as F_0 and intensity), while segmental features of speech deteriorate much more slowly. For example, the speech of most deaf patients prior to CI implantation has an abnormally high F_0. Once implanted, these patients showed an almost immediate reduction in F_0 towards normal levels (Leder et al., 1987). Subsequently, turning the implant off resulted in an elevation in F_0 to pre-implant levels.

Auditory feedback provides information not only about one's internal cues for regulating speech, but also provides feedback from the environment and about how others are responding to what was said. Additional supporting evidence for this concept comes from literature on the Lombard Effect (H. Lane & Tranel, 1971) and sidetone amplification studies (H. Lane & Tranel, 1971; H. L. Lane, Catania, & Stevens, 1961). The Lombard Effect shows that people increase the intensity (or loudness) of their voices (one acoustic aspect of prosody) to overcome noise in the environment. Similarly, sidetone amplification studies show that individuals will increase loudness due to reduction in sidetone volume (e.g. through headphones) and then voluntarily sustain their increased loudness. Data from the Lombard Effect and sidetone amplification studies, together with post-lingual deafness and cochlear implant research, demonstrate the importance of auditory feedback for prosody of speech. Thus, given the known prosodic abnormalities in speech of children with ASD (irregularities in pitch, tone, stress, or emotion) (Boucher, 2003; Rapin & Dunn, 2003;

Shriberg et al., 2001; Siegal & Blades, 2003), investigation of whether the audio-vocal regulatory system is functioning appropriately in ASD is warranted.

Measures of vocalizations in response to altered auditory feedback provide a view into the processing of auditory feedback and vocal motor control. A relatively new method, the pitch-shift reflex paradigm, has been developed for studying the relationship between auditory feedback and control of F_0. This technique allows one to quantitatively measure the audio-vocal system. In this technique, brief, unanticipated perturbations in voice pitch feedback are presented to subjects as they sustain vowels(Burnett et al., 1998; Hain et al., 2000), speak (Chen et al., 2007), or sing (Natke, Donath, & Kalveram, 2003). This paradigm reveals an automatic (or reflexive) mechanism for stabilizing voice F_0 by correcting for errors in voice F_0 production based on the auditory feedback.

Attempts to model audio-vocal control have suggested that auditory feedback acts as a negative feedback system to correct for errors in voice and F_0 production (Guenther, 2006; Guenther et al., 1998; Hain et al., 2000; Tourville et al., 2007). The Directions Into Velocities of Articulators (DIVA) model proposed by Guenther and colleagues provides a major theory for speech production that involves extensive interactions across many brain regions (Guenther, 2006; Guenther et al., 1998; Tourville et al., 2007). Further, they report that experimentation with speech begins early in development, as is evidenced by infant babbling. Hain and colleagues (2000) have proposed a response pathway for audio-vocal feedback whereby auditory input is compared with an internal or external referent to stabilize

voice F_0. Thus, it is proposed that vocal control involves a comparison of the voice auditory feedback with an internal (mental) representation of sound (i.e., referent memory) to achieve a goal (e.g., desired pitch or loudness). Moreover, effective communication relies on the ability to recognize when one needs to alter his or her speech in order to be better understood and to then adjust one's voice accordingly (H. Lane & Tranel, 1971). The concept of a "Theory of Mind" (Premack & Woodruff, 1978) enables a person to understand the point of view or mental state of others. Hence, having a Theory of Mind allows a person to recognize when he or she is not being understood (e.g., because of background noise) and there is a need to alter one's voice. This concept relates to the ideas expressed by the audio-vocal models of speech production in that the internal referent is the auditory memory and the goal is the desire to be understood. Because Theory of Mind is impaired in ASD, this inability may impede voice regulation during social interactions (McCann & Peppe, 2003; C. A. Miller, 2006).

Building upon what is known about the audio-vocal system and the problems regulating voice F_0 and atypical auditory processing of sound in ASD, the pitch-shift reflex was investigated in children with ASD. The aim of this study was to determine if children with ASD demonstrate normal or abnormal reflexive responses to pitch-shifted voice-feedback compared with age-matched typically-developing (TD) control children. We hypothesized that aberrant function in the audio-vocal system in children with ASD would result in abnormal voice production in response to auditory feedback manipulations in vocal pitch.

Methods

Participants

Study participants were recruited from community organizations and/or websites for families of children with ASD, as well as the "Chicago Parent Magazine." Participants included 19 TD children (11 males, 8 females) and 18 children with ASD (16 males, 2 females). For our purposes, the term ASD includes diagnoses of autism, Asperger Disorder, and Pervasive Developmental Disorder - Not Otherwise Specified (PDD-NOS). Children were required to have a formal diagnosis along the spectrum made by a child neurologist or psychologist and were actively monitored by their physicians and school professionals at regular intervals. In addition, diagnoses were supplemented by an internal parent questionnaire that detailed the child's developmental history, current symptoms, and functional level at time of entry into the study. Although the Autism Diagnostic Observation Schedule (ADOS) (Lord et al., 2000) and Autism Diagnostic Interview-Revised (ADI-R) (Lord et al., 1994) are the current research and academic standard for diagnosing ASD, many participants were diagnosed prior to the regular use of these instruments. Because these tests are not yet the standard for clinical diagnoses, we did not subject the children to additional testing and instead chose to accept their established clinical diagnoses for study inclusion. Parental reports of clinical diagnoses included autism (n=1), Asperger Disorder (n=6), PDD-NOS (n=1), and a combined diagnosis (e.g., PDD/Asperger Disorder; n=10).

Children were between the ages of 7-12 years (TD mean (SD)=10.00 (2.186); ASD =10.78 (1.865)) and chronologically age-matched across groups (one way ANOVA, $F_{(1,35)}$=1.349,

p=0.253). In the general population, the incidence of ASD in males is greater than in females. Because recruitment for this study was not restricted by gender and there were no known effects of gender on the pitch-shift reflex, children were not gender-matched. However, the two females in the ASD group were individually age-matched with two females in the TD group and analyses were performed to evaluate any gender differences. The children with ASD were all high-functioning and verbal. Although verbal ability and characteristics (i.e., echolalia, intonation abilities) were addressed in subject history questionnaires completed by parents, no formal evaluation of spontaneous speech was conducted. Thus, no quantitative measures of speech characteristics outside of the test paradigm were available for analysis. Other inclusion criteria for both groups were the absence of confounding neurological diagnoses (e.g., active seizure disorder, cerebral palsy), the presence of normal peripheral hearing determined by air threshold audiogram (thresholds $\leq$ 20 dB for pure tone octave frequencies 250-8000 Hz), and a full scale mental ability confidence interval score $\geq$ 80 (Wechsler Abbreviated Scale of Intelligence (WASI) (Woerner & Overstreet, 1999)) (Table 11). Age and gender were examined as possible covariates for the multivariate analysis of covariance (MANCOVA) of WASI scores. This preliminary MANCOVA indicated that they were not statistically significant; therefore subsequent multiple analysis of variance (MANOVA) tests were conducted without these covariates for mental ability comparisons. Although the children with ASD scored lower than the TD children on measures of full scale and verbal mental abilities ($F_{(1,35)}$=5.699, p=0.023 and $F_{(1,35)}$=9.011, p=0.005, respectively), the children with ASD tested well within the normal range on these measures and did not

differ on scores of performance mental ability ($F_{(1,35)}$=0.745, p=0.394). The normal scores provided confirmation that the children could comprehend the task requirements.

Behavioral tests

All behavioral testing was conducted in a quiet office by the experimenter who sat across a table from the child. Parents were invited to remain with their child if the child preferred, otherwise the parents sat in a lobby during testing. The WASI, which was an inclusion criterion test, is a test of mental ability (or IQ) and provides standardized scores of full scale mental ability, as well as verbal (vocabulary, similarities) and performance mental ability (block design, matrix reasoning). Additionally, the Clinical Evaluation of Language Fundamentals (CELF) (Semel et al., 2003) (Table 11) was administered to assess language ability and to provide standardized scores of core (overall), expressive, and receptive language abilities. Responses that required lengthy or specific answers were digitally recorded and transcribed for offline scoring after testing.

Pitch-shift reflex paradigm

The pitch-shift reflex was measured using procedures similar to those previously reported (Burnett et al., 1998; Larson et al., 2007). Briefly, the child sat comfortably in a chair while wearing Sennheiser HMD headphones with an attached Sennheiser microphone. The experimenter asked the child to produce a steady /a/ vocalization for periods of approximately five seconds, pause to take a breath, and then repeat. The experimenter demonstrated the task for the child and then the child practiced before the experiment began.

Because some of the children demonstrated reluctance to be in a confined sound booth for the testing, all subjects were tested in the main laboratory. The room was reserved strictly for the subject, parent, and tester, such that ambient background noise was equal across subjects. The low-level ambient noise in the room was not a problem because the headphones were the closed type, and there was the addition of 40 dB SPL pink masking noise to the auditory feedback to help reduce possible outside noises. Previous work has shown that pink masking noise does not alter the responses (Burnett et al., 1998). Acoustic calibrations made with a Brüel & Kjær sound level meter (model 2250) and in-ear microphones (model 4100) were used to set the computer display in calibrated units. Thus, to make sure subjects maintained a constant voice amplitude of about 75 dB SPL, the experimenter monitored the voice signal on the computer display and gave hand signals to the participant to raise or lower their voice amplitude as needed. Once it was apparent that the child understood the task and could comply with instructions, the experiment was initiated.

After the child began vocalizing, five randomly timed pitch-shifted stimuli (–100 cents (down one semitone), 200 ms duration) generated by a MIDI controlled Eventide Eclipse Harmonizer were incorporated into the voice signal in real-time and delivered through the headphones as feedback. Stimuli of 200 ms duration were used because they tend to elicit only reflexive responses as opposed to long durations that are more likely to trigger a voluntary response (Burnett et al., 1998). A stimulus magnitude of –100 cents was chosen because it is an established, standard stimulus and the most widely used for this type of study

(Bauer & Larson, 2003; Burnett et al., 1998; Hain et al., 2000); it is easily relatable to a music scale; and it is perceptible. Five stimuli were delivered with a 500-900 ms variable interstimulus interval within each 5-sec vocalization. This task was repeated approximately 16 times, totaling about 80 stimulus presentations. (The actual number of trials varied according to a given child's ability to hold his or her vocalization for five consecutive seconds.) The voice signal, a signal representing voice feedback, and TTL control pulses from the MIDI program were digitized using PowerLab (10 kHz per channel, 12 bit, 5 kHz anti-aliasing filter; AD Instruments) and recorded on a laboratory computer utilizing Chart software (AD Instruments, Colo. Springs, CO).

Analyses

Vocal responses were analyzed by first processing the voice and auditory feedback signals in PRAAT (Boersma & Weenink, 2004), which labeled each glottal cycle with a pulse. This pulse train was transferred to another program, Igor Pro (Wavemetrics, Inc., Lake Oswego, OR), where it was converted to an F_0 analog wave in which voltage corresponded to frequency. These F_0 signals were then converted to a cents scale using the following equation: cents=100 (39.86 log10 (f2/f1)) where f1 equals an arbitrary reference note at 195.997 Hz (G4) and f2 equals the voice signal in Hertz. The cents scale is a log scale that allows comparison of voice frequency across subjects who have different voice F_0 levels. Voice signals were aligned (in Igor Pro) with each stimulus onset TTL pulse on a computer monitor with a 200-ms pre- and 700-ms post-trigger window. The vocal responses were visually screened to remove trials with aberrant signals, and then an average response of

voice F_0 was generated from all the acceptable trials. Aberrant signals were usually the result of an error in the F_0 extraction in Praat, or a vocal interruption such as a cough. Averaged responses were produced separately for each child. The program then automatically detected changes in the voice F_0 waveform that exceeded 3 standard deviations (SD) of the pre-stimulus average, beginning at least 60 ms after the stimulus onset. The program measured the onset latency (time of this threshold crossing), magnitude of the response (greatest deviation in F_0 contour), and time of peak magnitude (difference between the latency at which the response magnitude is achieved and the onset latency) (Fig. 12). Individuals are more likely to produce a compensatory response to pitch-shifted feedback (i.e., a response in which the F_0 deflection is in the opposite direction to the stimulus). Less frequently, individuals will produce a "following" response (i.e., a response in which the F_0 deflection is in the same direction as the stimulus) (Burnett et al., 1998). The vocalizations were identified as compensatory or "following" based on the approximate morphology of the averaged response. Although the direction of response is not known to be a feature diagnostic of anything pathological, these data were separated into compensatory and "following" responses. In a separate analysis, variability in voice F_0 for each participant was measured by calculating the mean and SD of randomly chosen one-second voice samples of the F_0 contour in the absence of pitch perturbation stimuli. Local percent jitter was also calculated from the full duration of all vocalizations for each participant. Because of the multiple comparisons, a Bonferroni-adjusted alpha level of $p \leq 0.023$ (taking into account the inter-correlation among dependent variables; (Sankoh, Huque, & Dubey, 1997)) was determined necessary for a result to be deemed statistically significant.

Results

Pitch-shift reflex

Full scale IQ, age, and gender were examined as possible covariates with latency, time of peak magnitude, and magnitude of response. Preliminary analyses using a MANCOVA indicated that these measures were not statistically different; therefore subsequent statistical Mann-Whitney analyses were conducted without covariates.

Voice F_0 mean, variability (standard deviation), and local percent jitter did not vary between the groups in the study sample (Mann-Whitney, U=77, p=0.249; U=92, p=0.619; U=66, p=0.101, respectively). The TD children and children with ASD demonstrated a similar baseline; thus facilitating the interpretation of the following results.

Vocal responses to the perturbations were identified in all TD and ASD participants. Averaged responses across all children were based on an average of 65 trials (range 33-85). Sixteen of the TD children and 13 of the children with ASD produced compensatory responses, while 3 TD children and 5 children with ASD produced "following" responses. A Fischer's exact test was applied to these data to determine if there was any significance to the occurrence of compensatory versus "following" response patterns, and the two-tailed probability was not statistically significant (p=0.447). Given the low number of "following" responses in each group, meaningful statistics could not be evaluated for diagnostic comparisons of "following" responses. However, for descriptive purposes, group means and

standard deviations (SD) of "following" responses are as follows: onset latency (TD mean (SD)=0.16(0.061) sec; ASD=0.23(0.209) sec); time of peak magnitude (TD=0.05(0.015) sec; ASD=0.12(0.101) sec); and magnitude of the response (TD=7.49(1.391) cents; ASD=11.97(7.645) cents). Only compensatory responses are included in the subsequent data analyses.

In the group of children with ASD who produced compensatory responses (n=13), the diagnosis break-down included children with autism (n=1), Asperger Disorder (n=4), PDD-NOS (n=1), and a combined diagnosis (n=7). The TD and ASD groups were still age-matched (ANOVA, $F_{(1,27)}$=1.037, p=0.317; TD mean (SD)=10.06(2.265) years, ASD=10.85(1.772)). A MANOVA revealed no group difference in performance mental ability ($F_{(1,27)}$=0.845, p=0.366), whereas verbal mental ability did differ significantly ($F_{(1,27)}$=7.302, p=0.012) and full scale mental ability almost differed by the set criteria ($F_{(1,27)}$=5.003, p=0.034). However, the average mental ability scores were all within normal limits (Table 1). Mann-Whitney U tests revealed no main effect of diagnosis on any of the pitch-shift reflex measures, including onset latency (U=97, p=0.779), time of peak magnitude (U=88, p=0.503) and magnitude (U=103, p=0.983) (Table 12).

Language ability

A MANOVA revealed main effects of diagnosis on core and receptive language abilities (CELF; $F_{(1,27)}$=8.588, p=0.007 and $F_{(1,27)}$=12.245, p=0.002, respectively) such that children with ASD who produced compensatory responses had lower language ability scores than TD

children. However, children with ASD did not differ from TD children on measures of expressive language ability ($F_{(1,27)}$=1.362, p=0.253). Means and standard deviations are reported in Table 11.

Post-hoc analyses

Closer inspection of individual data revealed that the children with ASD showed two distinct compensatory response patterns; some children with ASD appeared to demonstrate a typical range of vocal F_0 modulations in response to perturbation, while others showed atypically large shifts in F_0 response magnitudes (Fig. 13). Because there are currently no normative data for children for this paradigm, and it is unknown how ASD may affect pitch-shift reflexes, compensatory responses were analyzed with respect to the mean TD magnitude (TD mean (SD)=22.11 (10.009) cents). There were no compensatory responses below –1.65 SD of the typical mean; therefore, separating out those responses above 1.65 SD captured the extreme 5% in the upper tail of the distribution. Response magnitudes that exceeded 1.65 SD of the TD mean magnitude were hence defined as atypical. The children with ASD were divided into two groups: those who were within 1.65 SD of the TD mean magnitude of voice F_0 responses to perturbation ("ASD-LOW," n=8) and those who had abnormally heightened voice F_0 responses ("ASD-HIGH," n=5). As is inherent in a normal distribution, one TD child also demonstrated a heightened response magnitude**,** but neither the inclusion nor exclusion of this child in the study altered the results. Because this child was without diagnosis, he was maintained in the TD group. Non-parametric Kruskal-Wallis and Mann-Whitney post-hoc tests were applied for subgroup analyses.

Group differences in WASI mental ability scores were examined between TD, ASD-LOW, and ASD-HIGH children (Table 11) and indicated no differences in performance mental ability ($H(2)=2.287$, $p=0.319$), verbal mental ability ($H(2)=5.21$, $p=0.074$) or full scale mental ability ($H(2)=4.825$, $p=0.09$). Age was re-explored with respect to the new groupings and no variance was observed between TD (10.06 (2.265) years), ASD-LOW (11.13 (1.808) years) and ASD-HIGH (10.06 (2.265) years) children ($H(2)=1.289$, $p=0.53$). Also, the ASD-HIGH children were not more likely to be of one specific spectrum diagnosis (ASD: autism: n=1, Asperger Disorder: n=1, PDD-NOS: n=1, combined diagnosis: n=2).

By definition, the ASD-HIGH children demonstrated statistically significant greater compensatory response magnitudes to pitch perturbation (Kruskal-Wallis test, $H(2)=14.764$, $p=0.001$). Follow-up Mann-Whitney tests showed that the ASD-HIGH group demonstrated larger responses than both the TD children ($U=17.0$, $p=0.001$) (Fig. 14) and the ASD-LOW children ($U=0.0$, $p=0.002$). However, the ASD-LOW group varied significantly from the TD group in terms of response magnitude ($U=26.0$, $p=0.019$) such that their mean magnitude was smaller than that of the TD group (with or without the TD child who exceeded the 1.65 SD cutoff). Onset latency did not differ between groups ($H(2)=6.507$, $p=0.039$). Time of peak magnitude also did not differ ($H(2)=2.258$, $p=0.323$; TD mean (SD)=0.22 (0.136) sec, ASD-LOW=0.24(0.208), ASD-HIGH=0.32 (0.155)). Means and standard deviations of response measures for each group are reported in Table 12.

Relationship to language

Kruskal-Wallis test results indicated a statistically significant group difference on receptive language ability ($H(2)=9.156$, $p=0.010$) and a near significant difference on core language ability ($H(2)=6.967$, $p=0.031$). However, expressive language ability did not differ between groups ($H(2)=4.825$, $p=0.090$). Mann-Whitney follow-up tests were conducted to examine differences in receptive language ability, and they showed a statistically significant group difference only between the TD and ASD-HIGH children ($U=5.5$, $p=0.002$). TD and ASD-LOW groups and ASD-HIGH and ASD-LOW groups did not vary significantly in receptive language ability ($U=37$, $p=0.106$ and $U=10.5$, $p=0.171$, respectively). For all CELF language measures (core, receptive, and expressive abilities), the TD children scored the highest, followed by the ASD-LOW children and then the ASD-HIGH children. Means and standard deviations of language measures for each group are reported in Table 11.

Irrespective of diagnosis, Pearson's correlations were calculated between the compensatory response measures (onset latency, time to peak and magnitude), WASI (full scale, verbal and performance mental abilities) and CELF (core, receptive, and expressive language abilities) behavioral measures. Correlations were considered significant if they both had p-values $\leq$ 0.05 and exceeded a value of +/-0.32; thus assuring that each meaningful relationship resulted in at least 10% shared variance between measures (Tabachnick & Fidell, 2007b). Response magnitude was significantly correlated with measures of core, receptive, and expressive language abilities ($r=-0.60$, $p=0.001$; $r=-0.55$, $p=0.002$; $r=-0.46$, $p=0.011$, respectively), such that decreased magnitude was related to higher language scores (Fig. 15).

Similarly, time of peak magnitude was also significantly correlated with core and receptive language abilities ($r=–0.37$, $p=0.048$ and $r=–0.44$, $p=0.017$, respectively), such that decreased time of peak magnitude was related to better language ability (Fig. 16). No statistically significant correlations were identified for measures of onset latency. When investigating diagnostic groups individually (data not shown), statistically significant correlations persisted between measures of response magnitude and core and receptive language indices and between time to peak and receptive language ability within the TD group and between response magnitude and core language index within the ASD group.

Discussion

This is the first study of which we are aware that reports pitch-shift reflex data on children in general and children with ASD, as well as the first to rigorously investigate the relationship to cognitive and language abilities. Since normative data for children in this age range do not exist, data from the TD children in this study represented the best control group. The children with ASD demonstrated two different types of responses to perturbation in auditory feedback; as a group, the ASD-LOW children (62%) responded with a smaller mean change in vocal F_0 in response to pitch-shifted auditory feedback than their TD counterparts, whereas 38% of the children with ASD showed larger response magnitudes. On an individual level, the children in the ASD-LOW group did not present with atypical response characteristics. It is only when looking at these 8 children as a group that they showed significantly smaller response magnitudes. However, what distinguishes the children in the ASD-HIGH group is that they showed abnormal response magnitudes on an individual level because their

responses were outside of 1.65 SD of the TD mean. Further, it is only the ASD-HIGH subgroup of children who showed significantly lower receptive language scores on the CELF than the TD children. Conversely, the ASD-LOW children did not differ on any language measure compared to TD children. These data indicate two potentially fundamentally different mechanisms of audio-vocal regulation in the ASD children of this study. One mechanism involves an audio-vocal system which is hypo-responsive or depressed, while the other mechanism may be a hyper-responsive audio-vocal system. Finally, across all children, correlations between pitch-shift reflex measures (time of peak magnitude and magnitude of the response) and behavioral language ability were identified, such that shorter time to peak and smaller response magnitude were indicative of better language abilities (as measured by the CELF).

One aim of this study was to identify a measure that may objectively characterize children on the spectrum. Not all children with ASD showed the same pattern of response, which is consistent with the known heterogeneity in ASD (Freitag, 2007; Tharpe et al., 2006). In this study, specific spectrum diagnosis alone (e.g., Asperger Disorder vs. PDD-NOS) did not account for the variation in pitch-shift reflexes. Provided the likelihood that the spectrum involves subpopulations with clinical features in common (Freitag, 2007), having a heterogeneous group of children with ASD showing two distinct types of effects is encouraging as a first step. Beyond correlating the pitch-shift reflex with available intelligence and language scores, other behavioral relationships were explored based on participant history reports. Because all of the children were receiving multiple kinds of

interventions (including speech therapy, occupational therapy, social skills groups, etc.), it was impossible to identify a common intervention that could account for differences in either language or voice F_0 regulatory abilities. An anecdotal observation by the experimenter was that nearly all of the children with ASD in this study demonstrated prosody production problems (including problems with volume, voice F_0, and intonation regulation). Further, parents often indicated either through personal communication with the experimenter or in response to study questionnaires that their child seemed to suffer from problems with both production and perception of prosody in speech. Consequently, the ASD-HIGH group did not distinguish itself from the ASD-LOW children as having a higher incidence of echolalia and flat intonation. Thus, the extent to which the pitch-shift reflex is related to echolalia or monotonicity could not be readily evaluated, particularly in the absence of formal measures of prosody production. Given the small sample sizes, these results speak to the need for future work in this area to distinguish between children with ASD who have smaller versus larger vocal responses and any accompanying behavioral or diagnostic correlates.

Currently available studies of vocal production in ASD rely on ratings of speech samples and offer only descriptions of the speech characteristics, rather than addressing why the speech is atypical(McCann & Peppe, 2003; Paul, Augustyn et al., 2005; Paul, Shriberg et al., 2005; Shriberg et al., 2001). Moreover, ceiling effects are commonly noted in behavioral measures of prosody in ASD (Paul, Augustyn et al., 2005; Paul, Shriberg et al., 2005). Data from the current study indicate the existence of objectively-measurable abnormalities in the auditory-vocal feedback loop in some children with ASD. In this study, mean F_0, low frequency F_0

variability (1-10 Hz; as in tremor) and cycle-to-cycle F_0 variability (voice jitter) did not differ between children with ASD and their TD counterparts. Thus, F_0 level and variability did not account for the differences in response to pitch perturbation (see (Liu & Larson, in press)). Therefore, it appears as though the children with ASD do not have an inherent deficit in the ability to sustain vocal F_0. Rather, it seems that children with ASD may have difficulty incorporating auditory feedback cues into vocal control mechanisms. The establishment of abnormalities in the audio-vocal feedback system is a first step for future investigations of prosody production and voice F_0 regulation in ASD. A recent study found differences in pitch range in children with ASD (Hubbard & Trauner, 2007). Since data on spontaneous speech characteristics (including voice F_0 range) were not available in the current study, exploring the relationship between natural speech and responses to audio-vocal feedback represents a logical next step in this line of research. Such studies would help to determine the extent to which echolalia, frequency range, or behavioral prosody may relate to audio-vocal reflexes in individual subjects.

A noteworthy model of audio-vocal interaction derives from birdsong literature (Margoliash, 2002; Prather et al., 2008). The process of crystallization of a song repertoire requires many steps, which may be homologous to vocal production in the human system (Marler & Sherman, 1983; Volman & Khanna, 1995). When a young bird first learns a song, it forms an auditory image of the sound. Once the image is solidified, the bird relies on auditory feedback, as well as feedback from the birds around it, to adjust its song. After modifications through the learning process, the song pattern crystallizes. Recent literature shows that in

response to auditory feedback manipulation at various times before, during, or after crystallization (a process referred to as "decrystallization"), the birdsong itself can be disrupted. It is encouraging to know that a song pattern specific to the repertoire of a given bird's species can be recovered after this disruption (Leonardo & Konishi, 1999). In addition, Prather and colleagues (Prather et al., 2008) have identified what appear to be audio-vocal mirror neurons which are active during listening and singing in the swamp sparrow. They further suggest that similar auditory-motor neurons may play a role in speech development in humans.

Drawing a parallel to birdsong development, a developing child must learn to produce speech patterns (Doupe & Kuhl, 1999). As a first step, a child forms auditory images of speech sounds. Using an internal model, the child then experiments with how to integrate the percept of a sound with the proper way to manipulate the vocal apparatus to produce the sound (babbling) (Ejiri, 1998; Guenther et al., 1998). If the percept of a sound is disrupted (at any level), then production of that sound would undoubtedly be affected. Furthermore, the production and regulation of voice F_0 during speech will have been "crystallized" with respect to this atypical representation. There are reports that in early development, children with autism show abnormal or absent babbling (Dawson et al., 2000; Gernsbacher, 2004; Iverson & Wozniak, 2007). Thus, one may hypothesize that the diminished experimentation with language through babble is related to the deficient audio-vocal feedback system.

The underlying neural circuitry in audio-vocal regulation involves many lower level nuclei, in addition to higher cortical processing. Because the latency of the pitch-shift reflex (130-200 ms) encompasses the time that it takes for a signal to travel from the midbrain to the motor cortex, both basic sensory encoding (lower level processing) and cortical encoding are likely involved. Although the present study paradigm precludes exact localization of the deficit in the audio-vocal system, some evidence for such localization emerges from work on vocal behavior and cortical activations in both humans (Houde, Nagarajan, Sekihara, & Merzenich, 2002) and non-human primates, such as the marmoset (Eliades & Wang, 2003). Based on results from an auditory feedback/magnetoencephalography study, Houde and colleagues (2002) suggested that cortical inhibition allows for online monitoring of speech output in comparison with expected vocalizations. Work by Eliades and Wang (2003) complement this theory; they showed in the marmoset that vocalization-induced inhibition in upper cortical layers begins before the onset of a vocalization, while excitation begins after the onset of vocalization, resulting in a cortical-cortical modulation. The working hypothesis suggested that inhibition allows the cortex to monitor auditory feedback of the self-produced vocal sounds, while excitation reflects responses to non-vocal environmental sounds. Furthermore, Eliades and Wang (2003) suggested corticofugal pathways may modulate (inhibition and excitation) cochlear and brainstem (specifically inferior collicular) responses to auditory vocal feedback. If the sensory auditory representation of the vocalization is precluded (on account of an atypical auditory neural pathway (Herbert & Kenet, 2007; Siegal & Blades, 2003)), then the cortex may not be receiving an appropriate signal to modulate the motor production of the sound. Alternatively, even if the sensory representation is accurate

and communicated to the cortex, there may be a disconnect between the cortical centers that modulate other cortical or lower level activity due to reduced inter-hemispheric or long-range connectivity (Baron-Cohen et al., 2005; Courchesne & Pierce, 2005). Given the known deficits in cortical processing of prosody in children with ASD (Erwin et al., 1991; Korpilahti et al., 2006; Kujala et al., 2005; Wang et al., 2001), one may speculate that the disruption in sensory-motor integration observed in the ASD-HIGH group in this study results from deficient cortical inhibition during vocalization via any of these plausible mechanisms.

The audio-vocal system relies on sensory-motor integration and individuals with ASD are often characterized as having deficits in this process (Iarocci & McDonald, 2006). Unfortunately, given the limitations of the current paradigm, it is impossible to know where exactly the disruption occurs in the auditory-motor pathway for vocal production. Even so, these data comprise the first representation of abnormalities in the pitch-shift reflex in children with ASD. These data show two patterns, ASD-LOW children who have diminished vocal responses and ASD-HIGH children who demonstrated larger responses. Due to their often flat or monotone vocal production, one might have predicted that children with ASD would not vary their voice F_0 in response to perturbation of auditory feedback at all and produce flat responses. These data show that the ASD-LOW group responds with a smaller change in voice F_0. This abnormality may either reflect a deficient automatic processing of the degree of pitch-shift stimulus, or it may reflect accurate recognition of the pitch-shift stimulus with a limited response by the vocal system possibly due to a behavioral abnormality (monotonicity). Conversely, individuals with ASD often self-report

hypersensitivity to sound (Kellerman et al., 2005; Khalfa et al., 2004; O'Neill & Jones, 1997). This auditory hypersensitivity may have contributed to the excessive disruption of the pitch-shift reflex mechanism observed in the ASD-HIGH group in this study. Either the auditory representation or vocal response may have higher gain. The ASD-HIGH children may be overcompensating for the pitch shift because of an initially heightened percept (in the auditory domain) with subsequent integration of sensory and motor systems required for voice F_0 production. Alternatively, the ASD-HIGH children may register the stimulus appropriately, but because they have relatively poor control over their vocal system, the result is a very large change in voice F_0. Regardless of sensitivity, abnormal auditory pathway function in general may be responsible for disrupted input into the initial stage of the auditory vocal motor system (Erwin et al. 1991; McClelland et al. 1992; Klin 1993; Maziade et al. 2000; Wang et al. 2001; Boddaert et al. 2003; Ceponiene et al. 2003; Jansson-Verkasalo et al. 2003; Rapin and Dunn 2003; Rosenhall et al. 2003; Boddaert et al. 2004; Gervais et al. 2004; Kasai et al. 2005; Kujala et al. 2005; Lepisto et al. 2005; Korpilahti et al. 2006; Lepisto et al. 2006; Tharpe et al. 2006). All of these possibilities warrant further study.

The robust relationship between audio-vocal production and language abilities is compelling. This relationship makes it possible to begin to consider measurement of the pitch-shift reflex as an early indicator of prosody-related language ability in children with ASD and to help identify candidates for more extensive and targeted language intervention. That is, the TD child who produced an abnormal pitch-shift response also demonstrated lower language abilities compared to his TD peers. Nevertheless, the possibility exists that these data are not

dichotomous in the ASD group, but instead represent a continuum of adolescent responses. Although developmental changes in the vocal tract and the role of auditory feedback have been modeled in adults (Callan, Kent, Guenther, & Vorperian, 2000), analogous data related to children are currently not available. Further, previous studies of the pitch-shift reflex have not evaluated language ability in adults. Understanding the maturation of the pitch-shift reflex and its relationship with language will help disentangle whether abnormal responses are indicative of ASD or poor language skills in general. Although it may be theorized that problems decoding acoustic aspects of speech may interfere with the learning of language skills, there is an admitted leap from perception and production to behavioral language abilities. Future studies are needed to explore the extent to which this relationship persists in larger samples, for both typically-developing and disordered children and adults.

In lieu of identifying a source of the deficit, it is encouraging to note that vocal production in response to auditory feedback may be malleable by training (Titze, 1994). Indeed, the neural encoding of pitch in the auditory system is malleable at both cortical (Jancke, Gaab, Wustenberg, Scheich, & Heinze, 2001) and subcortical levels (Krishnan et al., 2005; Krishnan et al., 2004; Musacchia et al., 2007; Wong et al., 2007; Xu et al., 2006). As demonstrated by training of singers, a person can learn to control voice F_0 range. With musical training, a child with ASD may learn how to appropriately gauge pitch in his or her own voice (i.e., integrating cues of vibration of vocal cords and pitch level) such that the perceptions of the individual's voice agree with the vocal productions. Remediation strategies involving vocal production and auditory feedback — either through speech or

music therapy — may address this problem in affected individuals. Furthermore, the pitch-shift reflex paradigm may be useful in monitoring effects of such therapies.

The pitch-shift response can reflect deficient and expert audio-vocal function. Patients with Parkinson's disorder, who have prosody production and voice F_0 deficits similar to individuals with ASD, also show abnormal pitch-shift reflexes consistent with what was observed in the ASD-HIGH group (Liu et al. "Vocal Responses to Loudness- and Pitch-shift Perturbations in Individuals with Parkinson's Disease" – Motor Conference abstract, 2008). On the other end of the continuum, audio-vocal experts (musicians) appear to have enhanced auditory-motor integration and can both detect pitch change better (Magne, Schon, & Besson, 2006) and are less affected by alterations in auditory feedback (Zatorre, Chen, & Penhune, 2007). Musicians exhibit a superior ability to ignore conflicting auditory feedback, while maintaining vocal output. The current findings, coupled with preliminary findings of abnormal magnitudes in patients with Parkinson's disorder patients and data indicating that musicians have a more finely tuned and accurate reflex, have significant theoretical implications. Additional investigations of altered auditory feedback and its effects on reciprocal pathways in the auditory-motor system are clearly needed to elucidate where deficits can be expected to occur. Identifying the actual mechanism will contribute greatly to the understanding of the continuum from deficient to expert auditory-vocal systems and the regulation and overall control of voice F_0.

The original impetus for this study was to link the observation that individuals with ASD often demonstrate abnormal perception and production of prosody with the audio-vocal feedback system. Identifying a difference in voice F_0 regulation between subsets of children with ASD and TD children on this audio-vocal feedback task was a first step and opens a new line of research. Further work is needed to determine the developmental time course of this feedback system and whether there are other characteristics that distinguish children with ASD with audio-vocal deficits from those in whom this feedback system appears to be intact. Future directions include 1) investigating other aspects of prosody (e.g., duration or rate); 2) implementing administration of the ADOS and ADI-R in order to confirm this phenomenon in a more homogenous group; and 3) determining how the audio-vocal response may align itself with specific social communication and behavioral deficits observed in children with ASD. The audio-vocal task is objective, non-invasive, reliable, and quickly measured (in less than fifteen minutes); it lends itself for use as an objective measure of one aspect of prosody deficits in ASD.

Acknowledgements

We would like to thank the children and their families who participated in this research. We would also like to thank Hanjun Liu, Steven Zecker, and Trent Nicol for their technical and statistical support. This work was supported by NIH R01 DC01510 and DC006243-02. We greatly appreciate the helpful comments from the anonymous reviewers. The authors declare that they have no competing financial interests.

CHAPTER VI: EFFECTS OF BACKGROUND NOISE ON CORTICAL ENCODING OF SPEECH IN AUTISM SPECTRUM DISORDERS

Submitted for Publication in *Journal of Autism and Developmental Disorders*

Nicole Russo[a,b], Steven Zecker[a], Barbara Trommer[b,c,d], Julia Chen[e], and Nina Kraus[a,b,f,g]

[a]The Roxelyn and Richard Pepper Department of Communication Sciences [b]Northwestern University Interdepartmental Neuroscience Program [c]Departments of Pediatrics and Neurology, Feinberg School of Medicine [d]Evanston Northwestern Healthcare Center for Neurodevelopmental Disabilities [e]Biological Sciences Department [f]Department of Neurobiology and Physiology [g]Department of Otolaryngology; Northwestern University, Evanston, IL, USA

Abstract

This study provides new evidence of deficient auditory cortical processing of speech in noise accompanying language impairment in autism spectrum disorders (ASD). Speech-evoked responses (~100 – 300 ms) in quiet and background noise and language abilities were evaluated in typically-developing (TD) children and children with ASD. ASD responses showed delayed latencies (both conditions) and reduced amplitudes (quiet) compared to TD responses. Expectedly, TD responses in noise were both delayed and reduced compared to quiet responses. However, no quiet-to-noise differences were found in ASD responses, presumably because responses in quiet were severely degraded. Finally, ASD responses in quiet mimicked TD responses in noise, indicating that children with ASD process speech in quiet only as well as TD children do in background noise.

Key Words: autism, cortical encoding, speech, background noise, children

Introduction

Autism spectrum disorders (ASD) are a class of developmental disorders that includes autism, Asperger's disorder, and Pervasive Developmental Disorder – Not Otherwise Specified (PDD-NOS). ASD is behaviorally defined by a triad of deficits: 1) language impairment with respect to social communication, 2) repetitive or stereotyped behaviors or interests, and 3) social isolation (Rapin & Dunn, 2003; Siegal & Blades, 2003; Tager-Flusberg & Caronna, 2007). The language impairment spans perceptual, productive, and physiological domains (Boucher, 2003; Herbert & Kenet, 2007; Kuhl et al., 2005; Shriberg et al., 2001; Siegal & Blades, 2003). One of the leading contributors to this language impairment is abnormal auditory processing (Rapin & Dunn, 2003; Siegal & Blades, 2003), which is similar to what is shown in other children with language-based learning problems(Cunningham et al., 2001; King et al., 2002; Warrier, Johnson, Hayes, Nicol, & Kraus, 2004; Wible et al., 2002, 2005). In one of the only known studies of speech perception in noise in children with ASD, Alcantara and colleagues (Alcantara et al., 2004) demonstrated elevated speech perception thresholds, poor temporal resolution and poor frequency selectivity.

Cortical Processing in ASD

Evaluation and characterization of cortical response timing and amplitude features can provide valuable insight into the sensory auditory processing deficits and language

impairment in ASD. Long latency auditory evoked potentials (P1, N1, P2, and N2) are known as exogenous cortical responses that are generated in primary or secondary auditory cortices (e.g., superior temporal cortex, planum temporale) and are elicited by the presence and physical features of an auditory stimulus (Ceponiene et al., 2001; Hall, 1992; Naatanen & Picton, 1987). Many of the earlier investigations of cortical processing in ASD focused on simple stimuli, such as tones (Bruneau, Bonnet-Brilhault, Gomot, Adrien, & Barthelemy, 2003; Bruneau, Roux, Adrien, & Barthelemy, 1999; Ferri et al., 2003; Gage, Siegel, Callen, & Roberts, 2003; Gage, Siegel, & Roberts, 2003; Lincoln, Courchesne, Harms, & Allen, 1995; Oades, Walker, Geffen, & Stern, 1988; Seri et al., 1999) and have investigated hemispheric differences or differential effects of encoding stimulus features such as frequency, duration, or volume. Results are mixed, though there is some indication of reversed asymmetries and immature response patterns. Further, much of the data on P1/N1/N2 responses have been extracted from studies focusing on endogenous responses involving oddball paradigms (mismatch-negativity, P3) or semantic processing (N4).

Due to the extensive literature on relationships between neural processing and speech perception in other populations (Cunningham, Nicol, Zecker, & Kraus, 2000; Cunningham et al., 2001; King et al., 2002; Warrier et al., 2004; Wible et al., 2002, 2005), as well as the influence of disrupted cortical organization and connectivity on the language impairment in ASD (Boddaert, Chabane, Gervais et al., 2004; Bruneau et al., 2003; Bruneau et al., 1999; Hardan, Muddasani, Vemulapalli, Keshavan, & Minshew, 2006; Herbert et al., 2002; Herbert et al., 2005; Jansson-Verkasalo et al., 2003; Just et al., 2004; Muller et al., 1999; Rojas,

Bawn, Benkers, Reite, & Rogers, 2002; Rojas, Camou, Reite, & Rogers, 2005), much of the current research is now focused on the cortical processing of speech. From studies specifically focusing on speech-evoked cortical potentials in *children* with ASD, there is some evidence for sensory processing deficits. Ceponiene and colleagues (Ceponiene et al., 2003) found a trend for reduced P1 amplitude in response to vowels in children with ASD compared to typical controls, whereas Jansson-Verkasalo and colleagues (2003) identified reduced N2 amplitude in response to consonant-vowel (CV) syllables in children with Asperger's disorder. In 2005, Lepisto and colleagues reported significant reductions only in P1 amplitude in response to vowels in children with autism and, then in 2006, Lepisto and colleagues reported no differences in children with Asperger's disorder.

Cortical Processing of Speech in Noise

Although as yet uninvestigated in ASD, analysis of cortical evoked responses to speech sounds in background noise has proven useful for investigating language impairment in other populations, such as children with language-based learning and reading problems (LP) (Banai et al., 2005; Cunningham et al., 2001; King et al., 2002; Warrier et al., 2004; Wible et al., 2002, 2005). These data stem from the literature indicating that background noise has a deleterious effect on both audibility and cortical processing of speech (Martin, Sigal, Kurtzberg, & Stapells, 1997; Whiting, Martin, & Stapells, 1998). Cunningham and colleagues (2001) reported more peak-to-trough amplitude reduction in children with LP than controls in cortical responses to the CV syllable /da/ presented in a background noise condition. In a study investigating effects of repetition and noise on cortical encoding of

speech, Wible and colleagues (2002) found that a subset of children with LP exhibited impaired response timing in noise that affected the correlation of responses under stresses of repetition. Using the same stimulus, several additional studies corroborated earlier results, including quiet-to-noise response correlation differences and latency differences in background noise in responses of children with LP compared to normal learning controls (Banai et al., 2005; King et al., 2002; Warrier et al., 2004; Wible et al., 2005). Moreover, auditory training has been shown to improve cortical responses to speech in background noise in children with LP (Hayes et al., 2003; Warrier et al., 2004).

Given the known abnormalities in speech-evoked cortical responses in quiet in ASD and the relationship between responses in noise to other language impaired populations, we hypothesized that children with ASD would demonstrate prolonged latencies and reduced magnitudes in response to speech stimuli presented in quiet and background noise compared with typically-developing (TD) children and that the between-group difference would be exacerbated in noise. Further, we predicted that the pattern of relationship between behavior and cortical responses would differ between groups. To test these hypotheses, we evaluated the effects of background noise on processing of the CV syllable /da/ in children with ASD and TD controls. Additionally, correlations between behavioral and neurophysiological measures were assessed.

Methods

The Institutional Review Board of Northwestern University approved all research and consent and assent were obtained both from the parent(s) or legal guardian(s) and the child.

Children were acclimated to the testing location and equipment prior to experimental data being collected. They were allowed to visit the laboratory and interact with the tester on multiple occasions. Some children brought electrodes home with them to better familiarize themselves with the neurophysiological procedure.

Participants

Participants included 16 verbal children with ASD (N= 14 boys, 2 girls) and 11 typically-developing children (TD, N = 7 boys, 4 girls). Age range was 7-13 years old and mean age (*SD*) did not differ between groups (TD, *M* (SD) = 9.82 (2.228) vs. ASD, *M* (SD) = 9.81 (1.682); independent two-tailed t-test; $t(25) = .008$, $p = .99$). Study participants were recruited from community and internet-based organizations for families of children with ASD. Children in the diagnostic group were required to have a formal diagnosis of one form of ASD made by a child neurologist or psychologist and to be actively monitored by their physicians and school professionals at regular intervals. Although the Autism Diagnostic Observation Schedule (ADOS; (Lord et al., 2000; Lord et al., 1989)) and Autism Diagnostic Interview-Revised (ADI-R; (Le Couteur et al., 1989; Lord et al., 1994)) are the current research and academic standard for diagnosing ASD, many participants were diagnosed prior to the regular use of these instruments. Because these tests are not yet the standard for clinical diagnoses, we did not subject the children to additional testing and instead chose to

accept their established clinical diagnoses for study inclusion. Parents were asked to supply the names of the examining professionals, their credentials, office location, date of initial evaluation and the specific diagnosis made. Diagnoses included autism (n=1), Asperger Disorder (n=5), and a combined diagnosis (e.g., Asperger Disorder/PDD-NOS; n=10). The diagnosis of ASD was supplemented by observations during testing such that included subjects were noted to have some or all of the following: reduced eye contact, lack of social or emotional reciprocity; perseverative behavior; restricted range of interests in spontaneous and directed conversation; repetitive use of language or idiosyncratic language; abnormal pitch, volume, and intonation; echolalia or scripted speech; and stereotyped body and hand movements. Diagnosis was also supplemented by an internal questionnaire that provided developmental history, a description of current symptoms, and functional level at time of entry into the study. Further inclusion criteria for both TD and ASD groups were 1) the absence of a confounding neurological diagnosis (e.g., active seizure disorder, cerebral palsy), 2) normal peripheral hearing as measured by air threshold pure-tone audiogram and click-evoked auditory brainstem responses and 3) a full-scale mental ability score ≥ 80.

Procedure

Hearing screening

On the first day of testing, children underwent a hearing threshold audiogram for bilateral peripheral hearing (≤ 20 dB HL) for octaves between 250 and 8000 Hz via an air conduction threshold audiogram on a Grason Stadler model GSI 61. Children wore insert earphones in each ear and were instructed to press a response button every time they heard a beep.

Cognitive and language testing

All behavioral testing took place in a quiet office with the child seated across a table from the test administrator. Full-scale mental ability was assessed by four subtests of the Wechsler Abbreviated Scale of Intelligence (WASI; Woerner & Overstreet, 1999). The WASI also provided scores of performance and verbal mental ability which were not part of inclusion criteria (Table 1; mean and standard deviations). Although the Clinical Evaluation of Language Fundamentals-4 (Semel et al., 2003) was not used as an inclusion criterion, it was administered to provide indices of core, expressive and receptive language abilities (Table 13).

Stimuli and data collection

All neurophysiological recordings took place in a sound attenuated chamber. During testing, the child sat comfortably in a recliner chair and watched a movie (DVD or VHS) of his or her choice. The movie soundtrack was presented in free field with the sound level set to < 40 dB SPL, allowing the child to hear the soundtrack via the unoccluded, non-test ear. To enhance compliance, children were accompanied by their parent(s) in the chamber. Children were permitted breaks during testing as needed.

All auditory stimuli were presented monaurally into the right ear through insert earphones (ER-3, Etymotic Research, Elk Grove Village, IL, USA). Responses were recorded via Ag-AgCl electrodes, with contact impedance of $\leq$5 kΩ. For click-evoked brainstem responses,

trials with artifacts exceeding 23.8 μV were rejected online. For cortical responses, trials with artifacts exceeding 100 μV were rejected online.

Click-evoked responses

A click stimulus, a 100 μs duration broadband square wave, was presented at a rate of 13/sec Hz and 80 dB SPL (Bio-logic Navigator Pro). Click-evoked responses were sampled at 24 kHz and were online bandpass filtered from 100-1500 Hz, 12 dB/octave. Three blocks of 1000 sweeps each were collected. Click-evoked auditory brainstem response wave V latencies were used to assess hearing.

Speech-evoked cortical responses

Auditory evoked potentials were recorded in response to a 40 ms speech syllable /da/ synthesized in Klatt (Klatt, 1980). Within this syllable, the voicing begins at 5 ms and the first 10 ms are bursted. The frequency components are as follows: F_0: 103-125 Hz, F_1: 220-720 Hz, F_2: 1700-1240 Hz, F_3: 2580-2500 Hz, F_4: 3600 (constant), F_5: 4500 (constant); F_2-F_5 comprise what is referred to here as high frequency (HF) information. The /da/ stimuli were presented with alternating polarity in order to minimize stimulus artifact and cochlear microphonic (Gorga, Abbas, & Worthington, 1985). The speech-evoked responses were collected in two different conditions, at a conversational speech level in quiet (80 dB SPL) and in background noise (+5 dB SNR).

For this part of the testing session, electrodes were placed centrally on the vertex (Cz), the contralateral earlobe (reference), the forehead (ground), and the superior canthus of the left eye (to monitor eye blinks). Speech stimuli were presented (Neuroscan Sound) with an inter-stimulus interval of 125 ms. Continuous white Gaussian noise was generated by a Biologic Navigator system and mixed with the /da/ stimulus in a Studiomaster mixer board to produce a signal to noise ration of 5 dB. Responses were recorded in Neuroscan 4.2 Acquire continuous mode with a sampling rate of 2000 Hz and a bandpass filter of 0.5-100 Hz (12 dB/octave), with a notch filter at 60 Hz, to isolate the frequencies that are most robustly encoded at the level of the cortex. Additionally, an online average was recorded simultaneously to monitor when approximately 1000 acceptable sweeps had been collected.

Data analyses

Data recorded from Cz were used for quantitative analyses. The first step of data reduction was to remove eye blink artifacts from the continuous EEG recording. This was done using a spatial filtering method implemented in Neuroscan 4.3 Edit. The eye-blink free file was then bandpass filtered from 1-40 Hz, and epoched using a 625 window (125 ms pre-stimulus period). An artifact rejection criterion of ± 65 μV was applied to the epoched file to remove sweeps containing large myogenic noise. The 100-300 ms time range of each of the remaining sweeps was then correlated with the corresponding time window of the *ad hoc* average of all artifact-free sweeps. Next, the sweeps were ranked according to how well they correlated with this average and the best 70% of correlated sweeps were used to create the final response average. Before performing statistical analyses, these final averages were pre-

stimulus baseline corrected to remove the DC drift. (Unless indicated otherwise, all data reduction was performed in Matlab 7.4.)

Based on a resemblance to responses in a previous study using subjects of similar age and a similar stimulus (Cunningham et al., 2001) and analysis of the grand average quiet response, the largest positive deflection (occurring approximately between 100-200 ms) was defined as the P1' response and the following negative trough (between 150-300 ms) was defined as N1'. For the background noise condition, individual responses were overlaid with the quiet responses and corresponding waveform morphology guided the choice of peak. An experienced peak picker manually picked all peaks and two additional peak pickers then confirmed these marks. All peak pickers were blind to subject diagnosis. Response measures included positive and negative peak latencies, peak-to-trough duration, amplitude and slope, signal-to-noise ratio, and quiet-to-noise response correlations.

Independent Student's t-tests (two-tailed) were used to evaluate group differences in click-evoked response latencies, mental and language abilities; the two-tailed result is reported because no differences were expected since all children met our inclusion criterion. Differences in cortical neurophysiology were first evaluated via a mixed design repeated measures analysis of variance (RMANOVA) to test the hypothesis that sensory encoding of speech in quiet and noise is disrupted in the cortex of children with ASD. Dependent variables included the cortical response measures listed above; the between-subjects factor was diagnosis; the within-subjects factor was condition (quiet versus noise). Based on the

prior data showing group differences with respect to cortical encoding of speech, our statistics were hypothesis-driven and thus, when appropriate, post-hoc analyses were conducted with one-tailed independent Students t-tests. To control for Type 1 errors during post hoc analyses, an adjusted alpha level of $p \leq 0.017$ was required for establishing significance. Additionally, paired t-tests within groups were conducted to evaluate whether children with ASD show the same effect of background noise as TD children. Levene's Test for Equality of Variances was applied to each statistical analysis and, when relevant, the reported p-values reflect corrections based on unequal variances. In order to discern any behavioral significance of cortical deficits, relationships between cortical response measures that differed between groups and cognitive and language abilities were evaluated via Pearson's correlations. Significant relationships were defined as r-values ≥ 0.35 and p-values ≤ 0.05. Note, for all statistical analyses involving quiet-to-noise response correlations, r-values were converted first to Fisher z-scores.

Age and sex considerations

Because of the broad age range, age was considered a variable in preliminary statistical analyses. There were no correlations between age and any of the dependent variables (Pearson's $r \leq 0.21$ $p \geq 0.191$). Further, due to the greater incidence of ASD in males versus females in the general population, the ASD group in this study included a majority of male participants. However, sex is not thought to affect cortical responses (Cunningham et al, 2001). Inclusion of age and sex as covariates indicated that they were not statistically significant, and as a result, they were not considered in subsequent analyses.

Results

Cognitive and Academic Testing

Children with ASD did not differ significantly from TD children on measures of full-scale ($t(25) = 1.13$, $p = 0.268$), verbal ($t(25) = 1.57$, $p = 0.129$) or performance ($t(25) = –0.26$, $p = 0.800$) mental ability, or expressive language ability ($t(25) = 1.31$, $p = 0.201$), but did differ from TD children on measures of core ($t(25) = 2.60$, $p = 0.015$) and receptive ($t(25) = 2.39$, $p = 0.025$) language abilities (see Table 13 for means and standard deviations).

Click-evoked Auditory Brainstem Responses

All children exhibited normal brainstem responses to click stimuli; there were no between group differences (ASD latency *M* (SD) = 5.58 (0.189) ms, TD *M* (SD) = 5.54 (0.163) ms; $t(25) = –0.57$, $p = 0.572$). As a combined group, the TD and ASD wave V latencies ranged from 5.15-5.79 ms, consistent with the previously reported normal range (Gorga et al., 1985; Hood, 1998; Jacobson, 1985).

Speech-evoked Cortical Responses

The mixed-design RMANOVA indicated significant between-group main effects on measures of P1' latency ($F_{(1,25)} = 7.37$, $p = 0.012$) and peak-to-trough amplitude ($F_{(1,25)} = 4.52$, $p = 0.044$). Follow-up protected independent t-tests (one-tailed) indicated delayed P1' latencies in both quiet and background noise ($t(25) = –2.27$, $p = 0.016$ and $t(25) = –2.51$, $p < 0.001$, respectively) and reduced peak-to-trough amplitudes in quiet ($t(25) = 2.57$, $p = 0.009$)

in children with ASD compared to TD children (Fig. 17). See Table 14 for means (SD) of all dependent variables.

Within group analyses (paired t-tests) indicated that noise adversely affected the TD response, while the ASD response was similar in quiet to noise ($t(15) \leq 1.81$, $p \geq 0.045$, all comparisons). Specifically, in background noise, the TD response demonstrated a delayed P1' latency ($t(10) = -2.73$, $p = 0.011$), reduced peak-to-trough duration ($t(10) = 2.5$, $p = 0.015$), and a reduced peak-to-trough amplitude ($t(10) = 2.48$, $p = 0.017$). As a final comparison, TD responses in noise were compared to ASD responses in quiet and revealed no significant differences in any of the dependent variables ($t(25) \leq 1.00$, $p \geq 0.163$, all comparisons; Fig. 18).

Correlations with Behavior

Correlations were computed between neurophysiological measures of P1' latency in quiet, P1' latency in background noise, peak-to-trough amplitude in quiet and behavioral cognitive and language abilities. Robust relationships between the latency measures existed only in the group of children with ASD. There was a significant relationship between P1' latency in quiet and background noise and verbal mental ability ($r = -0.57$, $p < 0.02$ and $r = -0.70$, $p = 0.003$ respectively), P1' latency in noise and full scale mental ability ($r = -0.65$ $p = 0.006$) and both core and receptive language ability ($r = -0.64$, $p = 0.01$ and $r = -0.61$, $p = 0.01$, respectively). For all of these relationships, earlier response latencies were related to better behavioral scores. Peak-to-trough amplitude did not relate to the behavioral scores.

Discussion

Summary

Children with ASD showed both timing and magnitude deficits in cortical processing of speech in quiet and timing deficits in background noise. These data represent the first demonstration of deficits in cortical encoding of speech in noise in ASD. As expected, TD children showed P1' latency, peak-to-trough duration and amplitude deficits when encoding speech in background noise. In contrast, children with ASD showed deficits in P1' latency and peak-to-trough amplitude compared to TD children, but no additive effect of background noise, such that in this study, children with ASD processed speech in both quiet and noise comparably to the manner in which TD children encode speech in noise. Further, correlations between significant response measures and behavior were only found in children with ASD. These data are consistent with recent findings of correlations between verbal mental ability and cortical evoked potentials in children with ASD (Salmond et al., 2007).

Our results show some variations from previously reported studies (Ceponiene et al., 2003; Jansson-Verkasalo et al., 2003; Lepisto et al., 2005; Lepisto et al., 2006), yet overall, they are consistent with converging evidence of cortical speech processing deficits in children with ASD. With the exception of Lepisto and colleagues' study (2006), reduced amplitudes have been reported in response to speech in quiet. The current study differed in that it examined a peak-to-trough amplitude rather than individual peak amplitudes. Because some children demonstrated N1' responses that were above baseline, peak-to-trough amplitudes were

reported here. Additionally, variations in results may be indicative of the different speech syllables and qualitatively different responses; our stimulus elicited a robust positive peak between 100-200 ms and a negativity occurring between 200-300 ms while others reported positive peaks in the 50-150 ms range and negative peaks in the 150- or 180-300 ms range. Also potentially accounting for differences is that both Ceponiene and colleagues (2003) and Lepisto and colleagues (Lepisto et al., 2005; Lepisto et al., 2006) used a vowel rather than a CV syllable. Children with autism in the study by Lepisto and colleagues (2005) differed significantly from the controls on measures of performance and verbal mental ability, whereas in this study, the children with ASD did not differ significantly from TD children on measures of mental ability. Finally, some studies focused only on specific subtypes of the ASD, i.e., children with Asperger's disorder (Jansson-Verkasalo et al., 2003; Lepisto et al., 2006) or autism (Ceponiene et al., 2003; Lepisto et al., 2005). Thus, the broad heterogeneity of children with ASD (Freitag, 2007; London, 2007; Salmond et al., 2007; Tager-Flusberg & Caronna, 2007) may account for the varying results. However, any of these disparities individually, or in combination, may have contributed to the differences with respect to peak amplitudes and latencies.

These results also indicate a different mechanism of deficit in children with ASD compared to other children with LP. Specifically, children with LP only showed deficits in background noise, such that poor timing in noise adversely affected quiet-to-noise response correlations (Cunningham et al., 2001; King et al., 2002; Warrier et al., 2004; Wible et al., 2002, 2005). Further, consistent with what the current study and prior literature indicate with respect to

encoding in background noise in TD responses, response amplitude was reduced in background noise in children with LP (Cunningham et al., 2001). This phenomenon was not observed in children with ASD. Instead, within the ASD group, responses were not significantly altered by background noise. Also different from children with LP, children with ASD showed impairments in cortical encoding of /da/ in *both* quiet and background noise conditions. Thus, children with ASD start out at a disadvantage for speech processing in quiet (significantly delayed P1' latency and reduced peak-to-trough amplitude) and maintain abnormalities processing speech in background noise (significantly delayed P1' latency) compared to TD children.

Cortical Abnormalities in ASD

Converging evidence from auditory evoked potentials and magnetic resonance imaging studies implicates abnormal differentiation of cortical areas important for language processing in ASD, which may explain, in part, the results of the current study (review in (Volkmar, Lord, Bailey, Schultz, & Klin, 2004)). Many studies reported reversed or absent asymmetry in the inferior frontal (Broca's area) and posterior superior temporal regions (Boddaert, Chabane, Gervais et al., 2004; Bruneau et al., 2003; Bruneau et al., 1999; Herbert et al., 2002; Jansson-Verkasalo et al., 2003), including the planum temporale (i.e., Wernicke's area) (Hardan et al., 2006; Rojas et al., 2002; Rojas et al., 2005) and higher-order association cortices (Herbert et al., 2005). Data also suggest reduced inter- and excessive intra-connectivity of the frontal cortex (Courchesne & Pierce, 2005; Just et al., 2004; Minshew & Williams, 2007; Wickelgren, 2005) and increased thickness in the temporal and

parietal lobes (Hardan, Jou, Keshavan, Varma, & Minshew, 2004). The prolonged latencies found in the current study may be indicative of aberrant connectivity such that sound cannot efficiently propagate the ascending auditory pathway, resulting in delayed latencies. Although this study did not show correlations with age, typically, cortical response latencies become earlier with maturation (Cunningham et al., 2000). Thus, these data may provide further support of an immature system in children with ASD (Gage, Siegel, & Roberts, 2003). Additionally, although one may suspect increased amplitude with a larger availability of neurons within the cortex, the observed reduced amplitude may be a result of poor coordination and decreased neural synchrony in response to speech. Further, neural noise in the cortex, associated with increased intra-connectivity and synaptic activity, may impede a robust stimulus-triggered response. Finally, another possibility is that reduced experience with language in ASD prevents normal development of auditory cortex.

Plasticity in the Cortex

Animal and human studies have shown that the organization of the auditory cortex and its projections are highly dependent on early exposure to sound (Buchwald, Guthrie, Schwafel, Erwin, & Van Lancker, 1994; Chang & Merzenich, 2003; de Villers-Sidani, Chang, Bao, & Merzenich, 2007; Keuroghlian & Knudsen, 2007; Nakahara, Zhang, & Merzenich, 2004; Zhang, Bao, & Merzenich, 2002). Much of our knowledge of critical and sensitive period development stems from studies in which auditory environments are disrupted. For example, pathways are reinforced with persistent exposure to specific sounds (de Villers-Sidani et al., 2007; Nakahara et al., 2004), whereas broad exposure to noise retards differentiation and

specialization (Chang & Merzenich, 2003; Zhang et al., 2002), and lack of exposure can preclude responsiveness to a sound (Buchwald et al., 1994) or results in reorganization of cortical areas (Fine, Finney, Boynton, & Dobkins, 2005; Kujala, Alho, & Naatanen, 2000). In adult animals, training- or experience-induced plasticity is dependent on behavioral significance of sound exposure (Recanzone, Schreiner, & Merzenich, 1993; Zhou & Merzenich, 2007). Further, the human auditory cortex is malleable with both cue enhancement and other commercial auditory training in both children with LP (Cunningham et al., 2001; Hayes et al., 2003; King et al., 2002; Warrier et al., 2004) and normal adults (K. Tremblay, Kraus, McGee, Ponton, & Otis, 2001; K. L. Tremblay & Kraus, 2002). Auditory training effectively improved timing and enhanced amplitudes of the cortical response and corresponded to behavioral improvements in speech perception.

Implications

This study provides new insight into the speech-related cortical processing deficits in quiet in ASD and implicates atypical cortical processing of speech in background noise. Further, these data reinforce the functional relationship between cortical speech processing and behavioral (cognitive and language) profiles in children with ASD. Given both the atypical cortical organization in ASD and the robustness of plasticity in the cortex, this paradigm may aid in the assessment of auditory remediation in ASD. Although children with a range of diagnosis of ASD were included, application of more precise tools for clinical classification of subjects, such as the Autism Diagnostic Observation Schedule (Lord et al., 2000; Lord et al., 1989) and Autism Diagnostic Interview-Revised (Le Couteur et al., 1989; Lord et al.,

1994), may improve our ability to characterize language deficits in children with ASD. Future studies should more vigorously investigate the relationship between neurophysiological effects of background noise and behavioral tests of speech perception in noise in children with ASD.

CHAPTER VII: DISCUSSION

Summary

These studies addressed whether speech-specific brainstem deficits occur in ASD, whether auditory processing deficits affect regulation of voice F_0, and whether brainstem deficits propagate to the cortex. The central hypothesis was that problems with speech and prosody (pitch) perception and production were due to disordered representation of speech in the brainstem and cortex. Relative to TD children, children with ASD exhibited speech encoding deficits in the absence of peripheral hearing deficits or abnormal processing of simple stimuli (clicks). Diagnostic group comparisons showed evidence for pervasive deficits in the brainstem and cortical transcription of the acoustic aspects of speech in quiet and noise in children with ASD, while a select few children with ASD lacked these speech-related central auditory processing deficits. Taken together, the results of these studies establish disordered brainstem and cortical processing of speech sounds (presented in quiet and background noise), identify pitch-specific deficits in subgroups of children with ASD (groups deficient either in pitch tracking or audio-vocal regulation of voice F_0), and confirm a relationship between such deficits and language abilities in ASD.

Although the groups varied slightly across studies, this comprehensive protocol provides a framework for objectively characterizing children with ASD based on tangible subgroup differences. This study successfully isolated individuals with impairment in pitch-related brainstem transcription (n=5) and individuals with abnormally large magnitudes in response to auditory perturbations (n=5). Consequently, pitch-specific neural processing deficits may be

related to behavioral deficits with prosody perception and production. Given the lack of sensitive measures of prosody perception in ASD (Paul, Augustyn et al., 2005), the application of pitch-specific paradigms (such as the brainstem pitch-tracking and pitch-shift reflex paradigms) may inform the design of new behavioral measures. Although as a group, brainstem transcription of speech is disrupted in children with ASD, not all of the children who demonstrated pitch-related brainstem transcription deficits also showed broader brainstem deficits (with respect to the more steady-state syllable /da/). Another subgroup of children with ASD was impaired in both brainstem and cortical processing of speech. Interestingly, deficits in voice F_0 regulation were largely independent of sensory auditory processing. Only one child with ASD who showed audio-vocal impairment also had a central auditory processing deficit. Brainstem transcription of the onset portion of the speech syllable /da/ was impaired in this child, whereas pitch-related acoustic cues were accurately transcribed. Thus, one may hypothesize that an inability to regulate voice F_0 is the result of disrupted interactions between the brainstem and sensory and motor cortices in some children with ASD. These findings suggest that while auditory processing may be intact, abnormal prosody control in ASD can result from aberrant higher order processing of sounds, deficient integration of auditory and motor responses, or uncoordinated motor output. Overall, it is important to consider that within each of these focused subgroups, there may only be two or three children showing a pattern of deficit. Even so, these data represent a first step in sub-categorizing speech-related neural processing deficits in ASD and they open up a promising line of research for identifying objective subgroups.

Corticofugal Modulation and the Neuroscience Perspective

Evoked potentials are useful for identifying timing and frequency encoding deficits and providing some information about the origin of these physiological deficits. As discussed in earlier chapters, contributions from recent neurobiological findings in ASD offer attractive candidates for explaining our findings, including deficits in brainstem development, cortico-cortical connectivity (Baron-Cohen et al., 2005; Courchesne & Pierce, 2005; Wickelgren, 2005) and myelination (maturation) (McClelland et al., 1992). Under-connectivity between cortices (especially a reduced concentration of fibers in the corpus callosum), over-connectivity within areas, and deficient corticofugal connections (e.g., pathways between the brainstem and cortex), and delayed maturation (myelination) may each play a role in the delayed latencies, reduced amplitudes, deficient pitch tracking in the brainstem, and aberrant control of the audio-vocal system. Irregular connectivity is one of the most replicable pieces of evidence in describing the neurological bases of ASD and these data support this theory. The male brain contains greater connectivity within regions and less connections via the corpus callosum and so this developmental pattern is one theoretical explanation for cortical development in ASD (Baron-Cohen, 2006; Baron-Cohen et al., 2005; Knickmeyer & Baron-Cohen, 2006). Researchers suggest that rather than chaotic organization, the disordered connectivity in ASD resembles the organization of an "extreme" form of male development.

Mirror Neuron System and Auditory-Vocal Feedback

Along with aberrant connectivity and development of brainstem and cortex, contributions of a disordered mirror neuron system (MNS) may account for some of the deficits in audio-vocal regulation of voice F_0 (de & Hamilton, 2008; Williams, Whiten, Suddendorf, & Perrett, 2001).

Beyond the auditory component (hearing a sound), people learn to produce sounds by watching others manipulate their vocal apparatus in order to make different sounds (Liberman & Mattingly, 1985). The relationship between the MNS and motor production of speech parallels how visualization of a behavior enhances performance (Gentili, Papaxanthis, & Pozzo, 2006; Yaguez et al., 1998). In many ways, activation of the MNS is a means of "practice" for the auditory-motor system pathways. Thus, successful communication requires functional connections between auditory, visual (imagery), and vocal motor systems. Due to the likely role of the MNS in communication and the reduced long-range connectivity in ASD, a disordered MNS is an attractive candidate for explaining audio-vocal deficits in ASD.

Further, as previously discussed, hearing sounds activates motor areas and there are both feedback and feed forward components to this audio-vocal system (Guenther et al., 1998; Zatorre et al., 2007). Extracellular recordings show the involvement of auditory-vocal mirror neurons in songbird vocal learning (Prather et al., 2008). If the auditory representation of intonation in speech is disrupted or absent, then the mirror neurons may be unable to associate a different sound with a different manipulation of the vocal apparatus. Thus, with a disrupted MNS and without the neural encoding of the vocal movements associated with producing different sounds – and potentially without the auditory percept of different articulation patterns to create different sounds – an individual with ASD cannot learn the motor manipulations for appropriate vocal production. Disrupted top-down modulation between sensory and motor areas may account for some of the breakdown of goal-related MNS activity.

Theory of Mind

Another key aspect explaining the disruption in the MNS is that "goal-directed" or "emulated" behavior is impaired amidst intact basic mimicry ability (de & Hamilton, 2008) and this aspect is associated with the concept of a disrupted Theory of Mind (ToM) in ASD (Oberman & Ramachandran, 2007). ToM is believed to develop by approximately four years of age and it involved being able to see things from another person's perspective and to understand ideas about people, things, or places when they are not readily visible or tangible (Premack & Woodruff, 1978; Steele, Joseph, & Tager-Flusberg, 2003). If a person is impaired in ToM, then they will likely be unable to make sense of acoustic social communication cues in speech (such as changes in pitch). Alternatively, if a person suffers from impaired central auditory representation of acoustic cues and cannot discriminate pitch contour or variations in prosody, then this can affect development of the ToM. Also, if a person is unable to relate variations in pitch inflection with the ability to convey different intentions or emotions, they will have no reason to learn to modulate pitch in their own voice and atypical production will ensue. Being able to extract what other people are thinking from how they say something is exceptionally difficult if one does not have access to variations in acoustic cues. The breakdown in ToM and communication may result from atypical processing of the acoustic aspects of speech, the inability to recognize the goal of vocalizations, and/or from not utilizing acoustic cues appropriately in vocal production (Siegal & Blades, 2003).

Plasticity in the Brainstem and Cortex

Extensive evidence from animal research and short- and long-term training paradigms in both children (e.g., King et al., 2002; Russo et al., 2005) and adults (e.g., Musacchia et al., 2007; Song et al., 2007; Tremblay & Kraus, 2002) shape what we know about sensitive periods, critical periods, and plasticity in the brainstem (e.g., Buchwald & Huang, 1975; Edeline & Weinberger, 1991a, 1991b, 1992; Huang & Buchwald, 1979, 1980; Yan & Suga, 1996) and auditory cortex (e.g., Buchwald, Guthrie, Schwafel, Erwin, & Van Lancker, 1994; Chang & Merzenich, 2003; de Villers-Sidani, Chang, Bao, & Merzenich, 2007; Keuroghlian & Knudsen, 2007; Nakahara, Zhang, & Merzenich, 2004; Zhang, Bao, & Merzenich, 2002). In humans, auditory system plasticity is often evaluated with behavioral training paradigms. The use of short-term auditory training programs in children with language-based learning disorders results in improvements in both auditory brainstem and cortical (Hayes et al., 2003; King et al., 2002; Russo et al., 2005) responses to speech in quiet and background noise. Given the nature of deficits in ASD, one may ask whether auditory training would benefit children with ASD. In an attempt to answer this question, a preliminary study investigating the benefits of a Fast For*Word* auditory-based language training program for children with ASD was initiated. Results show that speech-evoked brainstem, cortical, and audio-vocal responses are stable over time in control children with ASD and that such responses may serve as a barometer of training-related improvements in trained children with ASD (*unpublished data*). Not surprisingly, only a subset of trained children shows physiological changes. Corresponding improvements in behavioral language abilities are currently being explored.

Further, music and language skills are highly intertwined and given the malleability of the auditory and vocal systems, anything that taps into the auditory processing of complex sounds may result in improvements in language. Short-term linguistic training in adults (J. Song et al., 2007) and music training in children (Moreno & Besson, 2005) both result in improved pitch processing in speech. Multiple studies show that extensive music training enhances F_0 encoding in musicians versus non-musicians at the brainstem (Musacchia et al., 2007; Wong et al., 2007) and cortical levels (Schon et al., 2004). Schon and colleagues (2004) have also shown enhancements in the behavioral detection of pitch differences in language in individuals with long-term musical experience. Musicians are also less affected by conflicting auditory feedback (Zatorre et al., 2007). In the audio-vocal domain, musical ability and music therapy enhances control of prosodic features of speech (Stegemöller, Skoe, Nicol, Warrier, & Kraus, in press; Thompson, Schellenberg, & Husain, 2003). Individuals with ASD demonstrate significant delays when reaching major milestones, but access to the appropriate training paradigm has the potential to change the neural environment. Taken together, results of prior studies encourage the possibility for improvement and fine tuning of the auditory and vocal systems in this developmental disorder (ASD).

Practical Implications

Families are often "in the dark" about the prognosis for their child to improve and reach developmental milestones, as well as about which interventions or remediation programs would provide the best outcome for their child. Interestingly, the most significant correlations identified consistently across all experiments in this study were between physiology and core

(overall) and receptive language abilities as measured by the CELF. These data suggest that lower language abilities are related to poor neural processing of speech and poor audio-vocal control. Regardless of whether the poor neural processing is a precursor to the poor development of language skills or vice versa, such relationships speak to the benefit of using these objective tests in the identification of children with ASD who are more likely to present with lower language abilities.

Further, the results of these various studies may inform future work in remediation via commercially available auditory programs, as well as structured music therapy programs. Due to the robustness and stability of the responses described here, systematic comparisons of these responses prior to and following remediation may assist in the assessment of various modes of therapy. It is also conceivable that with the identification of children with ASD who have deficits in speech- or pitch-specific central auditory processing and/or vocal regulation, appropriate decisions about placement in an auditory training or music therapy programs would be possible. Families would have objective measures for determining whether certain forms of behavioral therapy were appropriate for their affected child.

Conclusion

Although the comprehensive experimental paradigm described here provides new evidence of speech-related neural processing deficits in auditory and vocal domains in children with ASD, there is still much to be learned about the etiology of the disorder and what contributes to the heterogeneity of the ASD population. Diagnostic advancements including the Autism

Diagnostic Observation Schedule (Lord et al., 2000; Lord et al., 1989) (behavioral observations of the child) and the Autism Diagnostic Interview (Le Couteur et al., 1989; Lord et al., 1997) (parent or caregiver report of a child's development) have been helpful in improving diagnoses. However, diagnostic guidelines still lack objective means of disambiguating the disorder (i.e., classifying where on the spectrum a child will fall) and determining the prognosis for children based their profile. Together, the results of these five experiments begin to inform the objective sub-classification of children on the autism spectrum and provide a means of identifying candidates for deficit-specific remediation programs. Keeping in mind that ASD is a developmental disorder and that the disordered systems are plastic, improvements in receptive and expressive language abilities appear to be within reach for children on the spectrum. Future studies investigating the breadth of speech-related neural processing deficits in larger cohorts of children with ASD, as well as in lower functioning study samples, and controlled studies of the efficacy of remediation programs will provide a significant contribution to the understanding of the neurological bases for the language impairment in ASD.

TABLES

Table 1
Brainstem response measures

	Timing	Magnitude
Transient responses	Peak latency (V, A, C, F) VA inter-peak interval VA inter-peak slope	Peak amplitude (V, A, C, F) VA inter-peak amplitude VA inter-peak area
Sustained responses	Correlations Stimulus-to-response Inter-response	RMS amplitude F_0 amplitude F_1 amplitude

The response measures indicated either timing or magnitude of the response. Slope is a composite component of timing and magnitude. RMS, root mean square; F_0, fundamental frequency; F_1, first formant frequencies.

Table 1. Brainstem Response Measures

Table 2
Normative values for discrete peak responses collected in both quiet (A) and background noise (B)

	Latency (ms)			Amplitude (μV)		
	n	Mean	SD	*n*	Mean	SD
A. Quiet						
Wave V	38	6.61	0.25	38	0.31	0.15
Wave A	38	7.51	0.34	38	−0.65	0.19
Peak C	38	17.69	0.48	38	−0.36	0.09
Peak F	36	39.73	0.61	36	−0.43	0.19
VA complex	38	0.89	0.19	38	0.97	0.28
VA complex area (μV × ms)				38	6.60	2.42
VA complex slope (μV/ms)	38	−1.21	0.37			
B. Noise						
Wave V	23	7.14	0.70	29	0.08	0.07
Wave A	22	8.38	0.78	28	−0.05	0.06
Peak C	36	18.00	0.88	36	−0.15	0.08
Peak F	31	40.01	1.04	32	−0.22	0.14
VA complex	22	1.26	0.47	28	0.13	0.10
VA complex area (μV × ms)				22	1.70	1.23
VA complex slope (μV/ms)	22	−0.13	0.05			

Due to the absence of certain peaks in individual waveforms, the number of subjects differs among measures.

Table 2. Normative values for discrete peak responses collected both in quiet (A) and background noise (B).

Table 3
Normative values for correlations of frequency-following responses collected in both quiet (A) and background noise (B)

	n	Mean	SD
A. Quiet correlations			
Stimulus-to-response (7–10 ms lead)	38	0.28	0.10
B. Noise correlations			
Stimulus-to-response (7–10 ms lead)	36	0.16	0.09
Inter-response (0–2 ms lead)	36	0.34	0.15

Correlations were conducted between the stimulus and response, as well as between responses collected in quiet and background noise.

Table 3. Normative values for correlations of frequency-following responses collected in both quiet (A) and background noise (B).

Table 4
Normative values for measures of the magnitude of the frequency-following response

	n	Mean	SD
A. Quiet			
F_0 amp	37	19.73	7.89
F_1 amp	36	8.46	2.23
RMS amp	38	2.32	0.72
B. Noise			
F_0 amp	26	13.56	6.89
F_1 amp	29	5.32	1.42
RMS amp	36	1.47	0.42

RMS, root mean square; F_0, fundamental frequency; F_1, first formant frequencies; amp, amplitude.

Table 4. Normative values for measures of the magnitude of the frequency-following response.

Table 5
Pearson's correlations among transient discrete peak measures

		Latency				Amplitude					Area	Slope
	WAVE	A	C	F	VA	V	A	C	F	VA	VA	VA
Latency	V	**0.89	**0.44	0.19	0.05	−0.25	0.18	0.17	0.17	−0.28	−0.26	−0.31
	A		0.30	0.24	**0.49	−0.27	0.04	0.26	0.21	−0.20	−0.01	** −0.47
	C			0.01	−0.14	−0.25	0.15	−0.10	0.00	−0.25	−0.30	−0.12
	F				0.20	0.28	0.06	*0.40	0.23	0.11	0.14	−0.02
	VA					−0.14	−0.27	0.23	0.16	0.10	**0.51	** −0.45
Amplitude	V						−0.15	−0.13	** −0.44	**0.70	**0.45	**0.67
	A							−0.02	0.28	** −0.81	** −0.82	** −0.58
	C								−0.01	−0.06	0.01	−0.17
	F									** −0.47	−0.31	**0.48
	VA										**0.85	**0.83
Area	VA											**0.46

A relationship was considered strong if $r = \pm 0.30$ and $P \leq 0.05$. Transient measures were highly correlated. $^{**}P \leq 0.01$; $^{*}P \leq 0.05$.

Table 5. Pearson's correlations among discrete peak measures.

Table 6
Pearson's correlations among sustained measures

	RMS amp	F_0 amp	F_1 amp
S−R corr	0.04	0.33	0.33
RMS amp		0.25	−0.03
F_0 amp			0.32

Relationships among sustained measures were less prevalent. S−R corr, stimulus-to-response correlation; RMS, root mean square; F_0, fundamental frequency; F_1, first formant frequencies; amp, amplitude.

Table 6. Pearson's correlations among sustained measures.

Table 7
Pearson's correlations between transient and sustained measures

	Latency					Amplitude					Area	Slope
	V	A	C	F	VA	V	A	C	F	VA	VA	VA
S–R corr	−0.06	−0.06	−0.06	0.33	0.02	0.02	−0.12	0.31	−0.20	0.11	0.14	0.12
RMS amp	0.30	0.15	**0.51	0.02	−0.23	−0.23	0.13	0.07	0.06	−0.22	−0.31	−0.06
F_0 amp	−0.01	−0.08	0.16	0.13	−0.16	0.17	0.19	−0.07	* −0.39	0.25	0.08	0.32
F_1 amp	** −0.45	** −0.50	−0.05	0.11	−0.25	**0.59	−0.23	0.01	** −0.53	**0.53	0.29	**0.61

Although some relationships exist between these measures, they are also largely independent response measures. $**P \leq 0.01$; $*P \leq 0.05$; RMS, root mean square; F_0, fundamental frequency; F_1, first formant frequencies; amp, amplitude.

Table 7. Pearson's correlations between transient and sustained measures.

Mental and language ability scores

Test battery	Subtest	TD (n=18) Mean (SD)	ASD (n=21) Mean (SD)
WASI (mental ability)	Full scale	117.78 (11.909)	107.33 (14.62)
	Verbal	115.89 (2.683)	101.95 (15.045)
	Performance	115.39 (11.62)	110.9 (13.707)
CELF (language ability)	Core	112.39 (9.388)	98.71 (20.219)
	Expressive	113.22 (12.638)	102.57 (22.171)
	Receptive	113.06 (7.712)	96.71 (19.233)

Table 8. Mental and language ability scores. Mental ability was assessed using the Wechsler Abbreviated Scale of Intelligence (WASI); it provided full-scale, verbal and performance intelligence scores. Although significantly lower than TD children on measures of full-scale (overall) and verbal measures, the children with ASD demonstrated scores well within the normal range on all three components. Language ability was assessed using the Clinical Evaluation of Language Fundamentals – 4th Edition (CELF); this test provided indices of core, expressive, and receptive language abilities. For all indices of language ability, the children with ASD scored significantly lower than TD children, but within the normal range. Mean scores (standard deviations) are reported.

Table 9

Significant speech-evoked auditory brainstem response measures

	TD (n=18)	**ASD (n=21)**
	Mean (SD)	**Mean (SD)**
Quiet response measures		
Wave V latency (ms)	6.54 (0.174)	6.73 (0.267)
Wave A latency (ms)	7.48 (0.232)	7.85 (0.412)
Onset response VA duration (ms)	0.94 (0.168)	1.13 (0.323)
Wave D latency (ms)	22.38 (0.433)	22.77 (0.549)
Wave F latency (ms)	39.25 (0.27)	39.54 (0.4)
Noise response measures		
Wave F amplitude (μV)	-0.11 (0.079)	-0.05 (0.096)
Stimulus-to-response-in-background-noise lag (ms)	8.85 (0.961)	8.17 (1.112)
Stimulus-to-response-in-background-noise correlation (r-value)	0.21 (0.088)	0.14 (0.108)
Quiet-to-noise inter-response correlation (entire response) (r-value)	0.4 (0.185)	0.27 (0.166)
Quiet-to-noise inter-response correlation (11-40 ms range) (r-value)	0.49 (0.206)	0.36 (0.23)
Composite response measures		
Onset synchrony in quiet	0.00 (2.359)	-3.84 (4.192)
Transient responses in quiet	0.00 (3.49)	-5.84 (5.714)
Phase locking in quiet	0.00 (1.423)	-1.99 (2.349)
Neural synchrony in noise	0.00 (2.603)	-3.17 (3.422)

Table 9. Significant speech-evoked auditory brainstem response measures. Significant differences (p <0.05) between TD children and children with ASD were identified for several speech-evoked auditory brainstem response measures. Mean values (standard deviations) are reported.

	TD (n=21)		ASD IN (n=16)		ASD OUT (n=5)	
	Mean	SD	Mean	SD	Mean	SD
f_0 Frequency Error (Hz)	8.52	2.201	9.16	2.216	14.10	1.998
f_0 Slope Error (Hz/sec)	35	20.2	35	23.9	120	45.6
f_0 Pitch Strength (r-value)	0.40	0.198	0.32	0.154	0.23	0.169
H2 Frequency Error (Hz)	13.43	2.071	14.19	1.893	17.86	1.534
H2 Slope Error (Hz/sec)	66	48.0	56	28.0	64	23.0

Table 10. Means and standard deviations (SD) for individual pitch-tracking measures for TD, ASD IN and ASD OUT groups. Note that the means of the TD and ASD IN group were similar, while the ASD OUT group (as determined by the Composite Score) had pervasive deficits.

		WASI Mental Ability Scores			CELF Language Indices		
		Full Scale	Verbal	Performance	Core	Expressive	Receptive
TD (n=19)	Mean	118.95	117.26	116.42	114.11	113.53	113.58
	SD	10.972	12.301	11.725	9.492	11.197	7.89
ASD (n=18)	Mean	109.33	103.56	113.11	101.94	106.89	99.78
	SD	13.521	15.382	11.585	16.148	18.626	15.318
TD (n=16)	Mean	117.75	116.25	115.56	113.56	112.06	113.25
	SD	10.933	12.593	12.372	10.046	11.186	7.937
ASD (n=13)	Mean	107.00	101.31	111.31	98.85	97.46	105.08
	SD	14.944	17.182	12.419	16.757	15.804	20.540
ASD-LOW (n=8)	Mean	107.63	101.68	112.63	103.38	106.88	102.13
	SD	12.794	16.677	8.684	15.611	16.357	16.111
ASD-HIGH (n=5)	Mean	106.00	101.60	109.20	91.60	102.2	90.00
	SD	19.532	19.970	17.936	17.587	27.941	13.491

Table 11. Behavioral test scores. Group means and standard deviations for scores on tests of mental (WASI) and language (CELF) abilities are reported for comprehensive TD and ASD groups. Subsequent analyses were restricted to children who produced compensatory responses. Means and standard deviations are reported for these groups as well (below the double line). Finally, post hoc analyses of compensatory responses resulted in a sub-division of the children in the ASD group into ASD-LOW and ASD-HIGH groups; their behavioral scores are also reported.

		Pitch-shift reflex measures		
		Onset Latency (sec)	Time to peak (sec)	Magnitude (cents)
TD (n=16)	Mean	0.24	0.22	22.11
	SD	0.140	0.136	10.009
ASD (n=13)	Mean	0.21	0.27	28.65
	SD	0.091	0.186	23.059
ASD-LOW (n=8)	Mean	0.25	0.24	13.19
	SD	0.084	0.208	4.715
ASD-HIGH (n=5)	Mean	0.13	0.32	53.38
	SD	0.031	0.155	17.722

Table 12. Pitch-shift reflex compensatory response measures. Group means and standard deviations for compensatory response onset latency (sec), time of peak magnitude (sec), and magnitude (cents) are shown for TD and ASD groups. Means and standard deviations are also reported for the ASD-LOW and ASD-HIGH subgroups below the double line.

Mental and language ability scores

Test battery	Subtest	TD Mean (SD)	ASD Mean (SD)
WASI (mental ability)	Full scale	116.09 (11.193)	110.25(12.322)
	Verbal	113.73 (13.016)	104.94 (15.097)
	Performance	112.00 (10.835)	113.19 (12.475)
CELF (language ability)	Core*	115.73 (7.309)	103.94 (15.062)
	Expressive	115.91 (10.348)	108.00 (17.944)
	Receptive*	112.36 (8.477)	100.19 (15.276)

Table 13. Mean (SD) of cognitive and language scores are reported for both typically-developing children (TD; n=11) and children with autism spectrum disorders (ASD; n=16). Full scale, verbal, and performance mental ability were measured by the Wechsler Abbreviated Scale of Intelligence and core, expressive, and receptive abilities were assessed using the Clinical Evaluation of Language Fundamentals (4th Edition). Children with ASD differed significantly from TD children on measures of core and receptive language ability (*$p \leq 0.017$). However, they demonstrated similar ability on measures of full scale, verbal, and performance mental ability and expressive language ability.

Cortical response measures

	Quiet		Noise	
	TD	**ASD**	**TD**	**ASD**
	Mean (SD)	**Mean (SD)**	**Mean (SD)**	**Mean (SD)**
P1' Latency (ms)*	140.14 (20.625)	159.69 (22.809)	151.91 (14.193)	169.91 (20.571)
N1' Latency (ms)	235.41 (21.432)	244.91 (30.480)	236.41 (26.693)	247.41 (32.117)
Peak-to-trough duration (ms)	95.27 (22.202)	85.22 (19.211)	84.50 (23.671)	77.50 (29.051)
Peak-to-trough amplitude (μV)*	6.56 (1.894)	4.66 (1.899)	5.22 (2.109)	4.13 (1.961)
Slope (μV/ms)	-15.63 (5.725)	-20.10 (6.070)	-18.16 (7.765)	-22.47 (10.652)
Signal-to-noise ratio RMS	9.13 (4.574)	10.67 (5.504)	8.97 (6.147)	8.23 (3.376)
Quiet-to-noise correlation (r-value)	--	--	0.82 (0.467)	0.54 (0.690)

Table 14. Cortical response measures. Mean (SD) of cortical response measures in quiet and background noise conditions are reported for both TD children (n=11) and children with ASD (n=16). Responses of children with ASD demonstrated significantly prolonged P1' latencies in quiet and background noise, as well as reduced peak-to-trough amplitudes in quiet compared to responses of TD children (*$p \leq 0.017$).

FIGURES

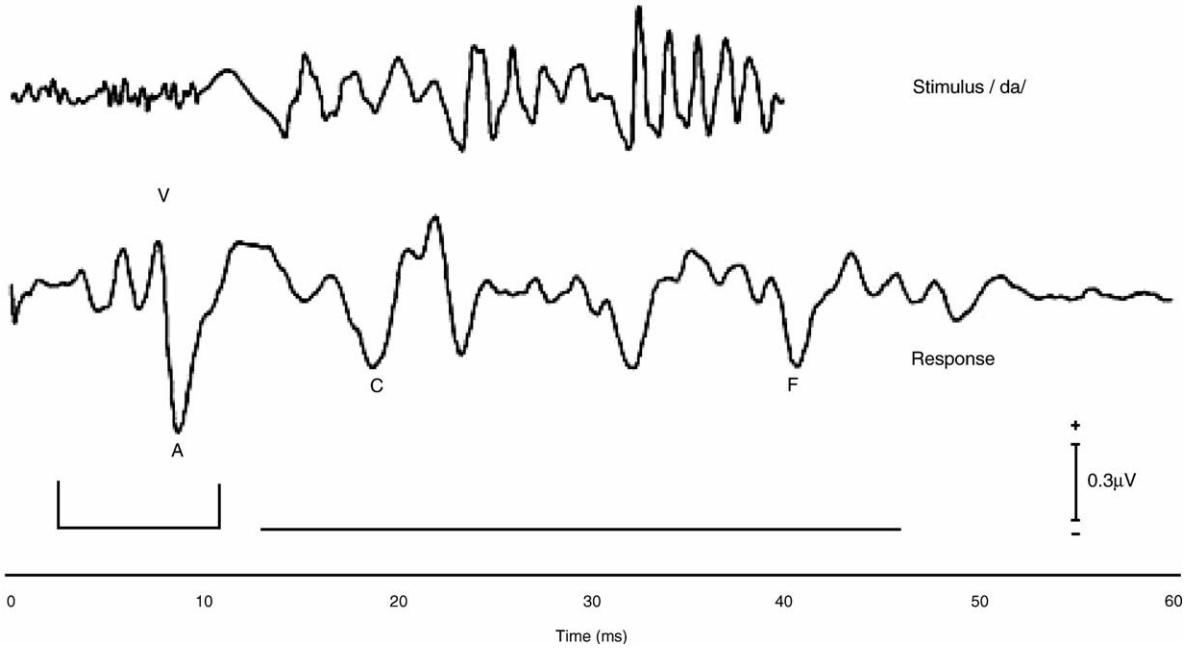

Fig. 1. Stimulus waveform (top) and grand average brainstem response in quiet (bottom; n=38). Three reliable negative peaks, waves A, C, and F, follow wave V. The onset response is bracketed, while the region containing the frequency-following response is indicated with a horizontal line.

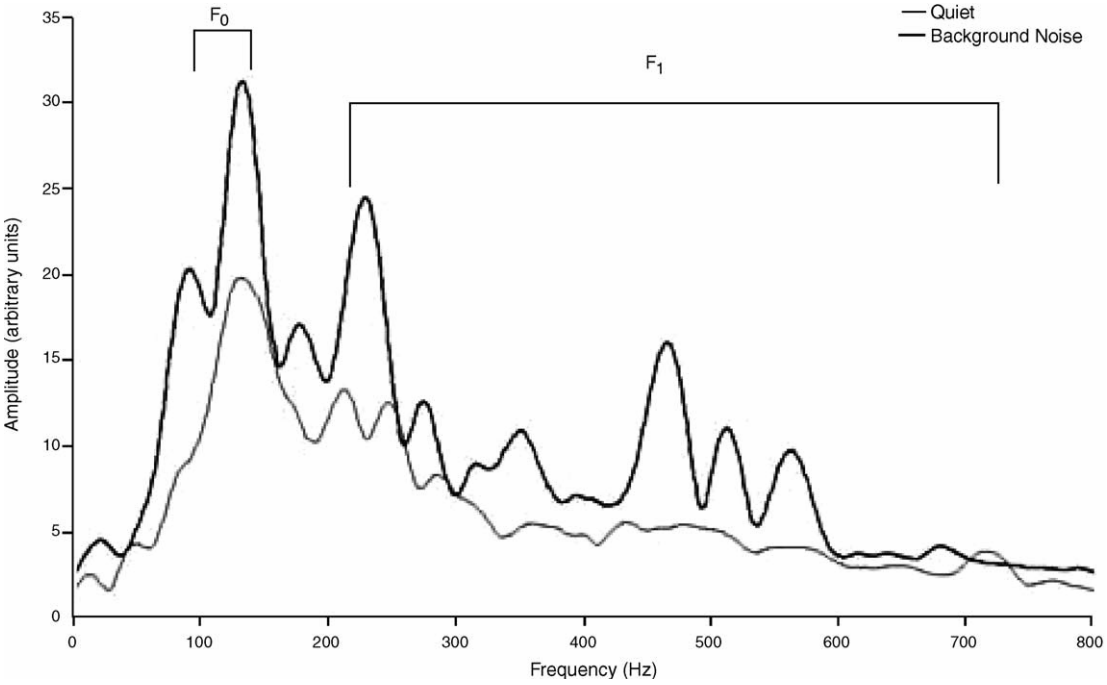

Fig. 2. Grand average frequency content in responses collected in quiet (n=36) and background noise (n=22). Analysis of the responses indicated that only the fundamental frequency and first formant (F0=103–121 Hz; F1=220–720 Hz) were measurable, whereas the higher frequency formants were not above the noise floor.

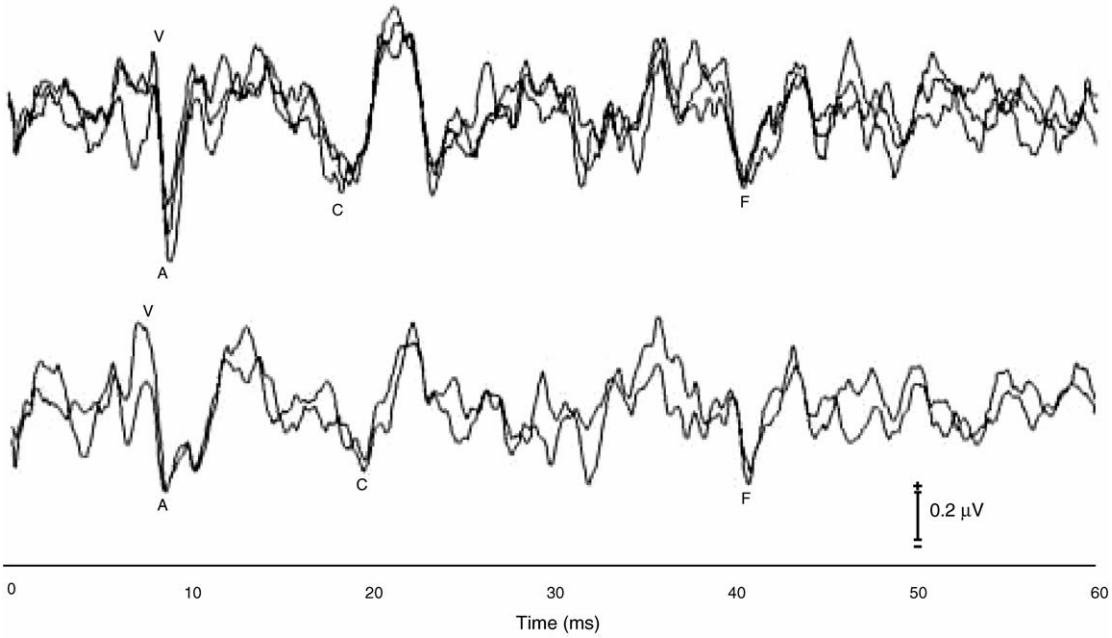

Fig. 3. Top: intra-subject, intra-test session reliability. Illustrated are three 1000-sweep subaverages that contributed to the final 3000-sweep response obtained for a representative subject. Bottom: intra-subject, inter-test session reliability. In another subject, two 3000-sweep averages were obtained on different test dates.

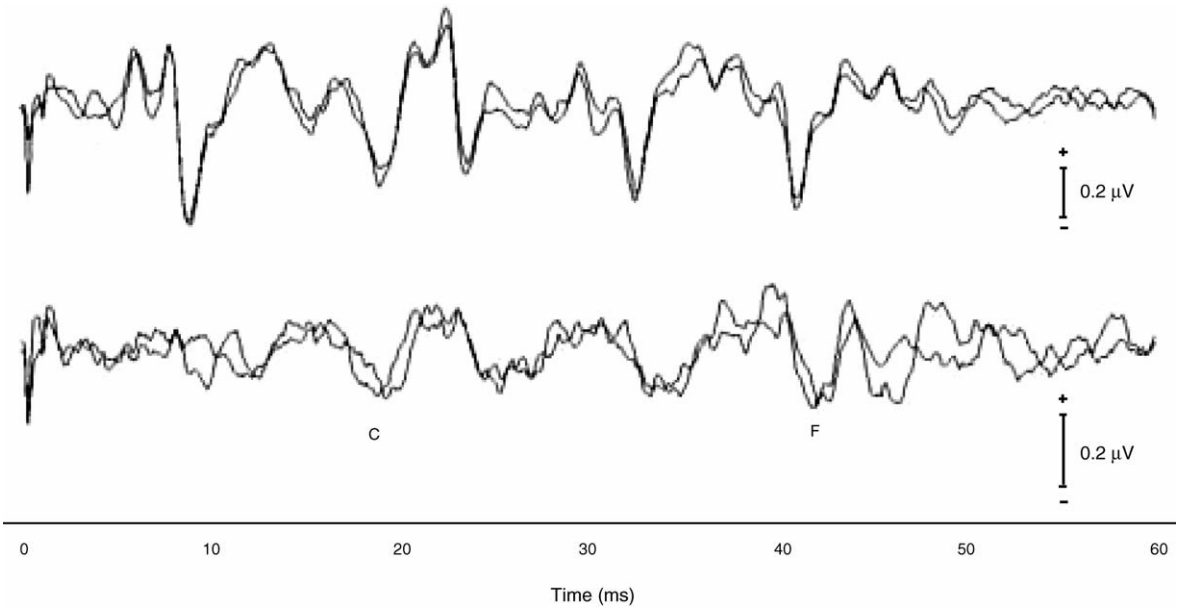

Fig. 4. Test–retest reliability. Grand average response waveforms collected in quiet (top) and background noise (bottom) at two different test sessions (n=8): Background noise effectively disrupts the onset response, while the frequency-following response remains intact.

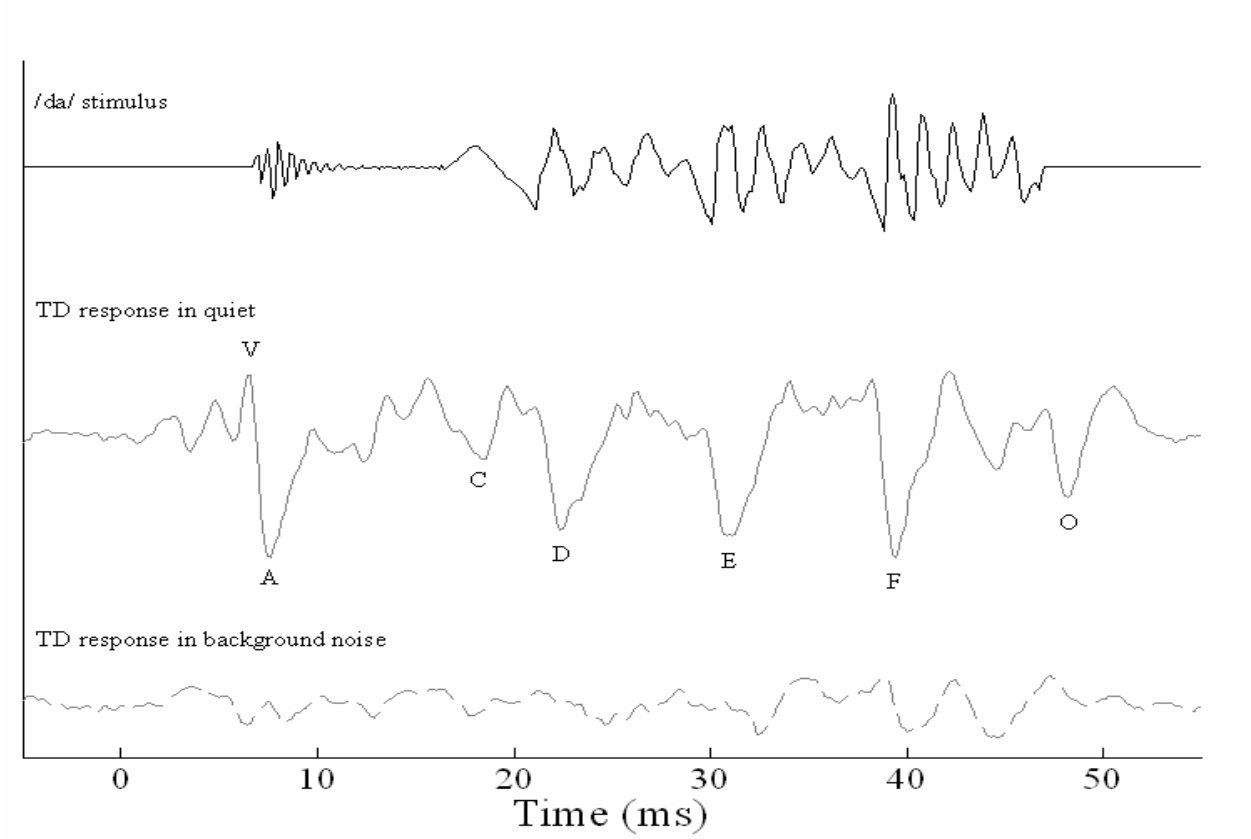

Fig. 5. Speech stimulus /da/ and TD grand average response in quiet and background noise conditions. The /da/ has an onset burst followed by a transition to the periodic vowel portion. The stimulus waveform is shifted ~7 ms to compensate for neural lag in the response. Brainstem responses to /da/ were robust in quiet, reflecting stimulus features with great precision. Waves V and A reflect the onset of the /da/ stimulus, wave C represents the transition to the periodic portion, waves D, E, and F comprise the frequency-following response, and wave O signals the offset of the response. The wavelengths between waves D-E and E-F correspond to the fundamental frequency (F0, pitch) of the stimulus, while F1 and higher frequency components are encoded in the smaller peaks between the dominant F0 waves. In background noise (dashed line), many of the transient response peaks are abolished, while sustained activity (frequency-following response) and waves F and O persisted.

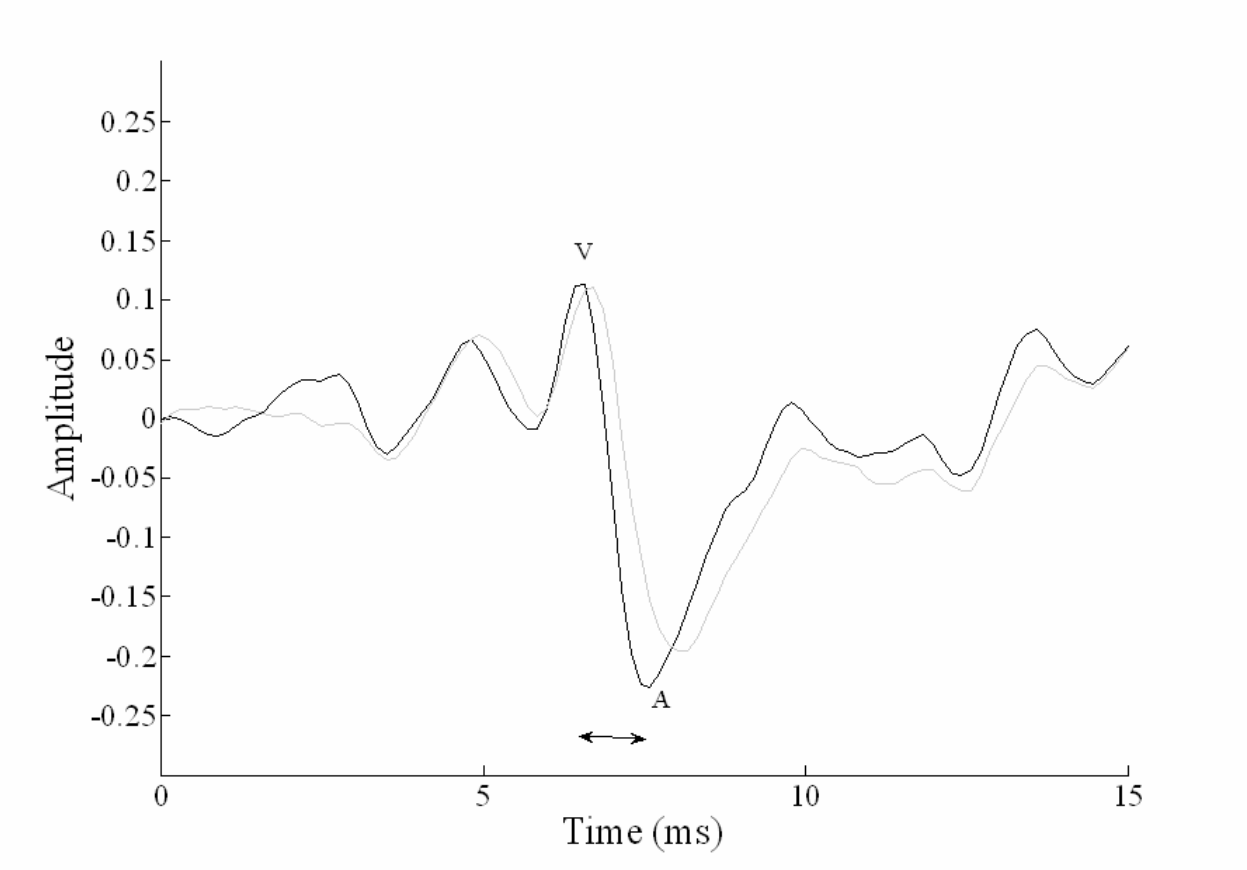

Fig. 6. Comparison of grand average onset responses to /da/ in quiet in TD children (n=18; black line) and children with ASD (n=21; gray line). The neural response to the onset of speech sounds was less synchronous in children with ASD (gray) as compared to TD children (black). Notably the onset response in children with ASD showed significant delays in waves V and A, and also a longer interpeak interval (horizontal arrow).

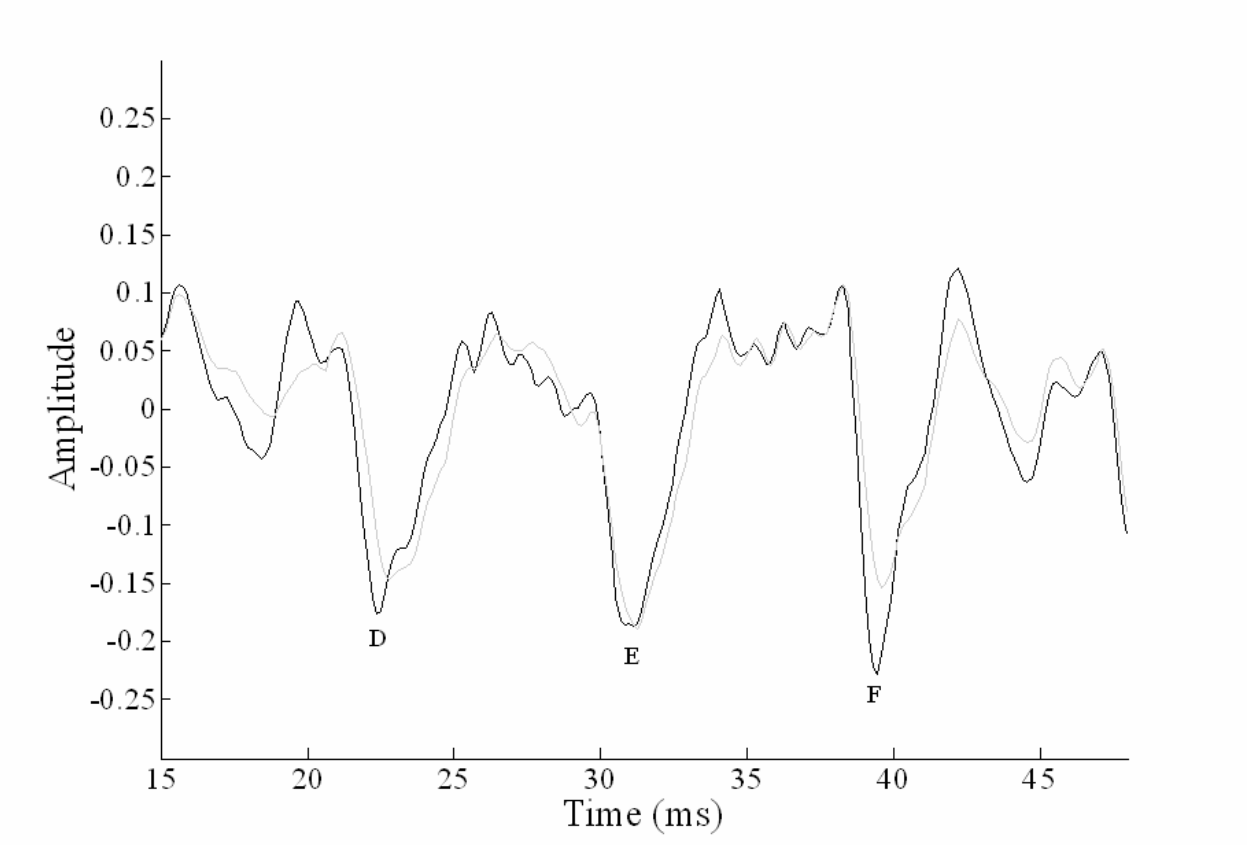

Fig. 7. Comparison of grand average frequency-following responses to /da/ in quiet in TD children (black line) and children with ASD (gray line). The frequency-following response in children with ASD showed significant delays in peaks D and F; peak F was also reduced in amplitude, demonstrating reduced phase locking in brainstems of children with ASD.

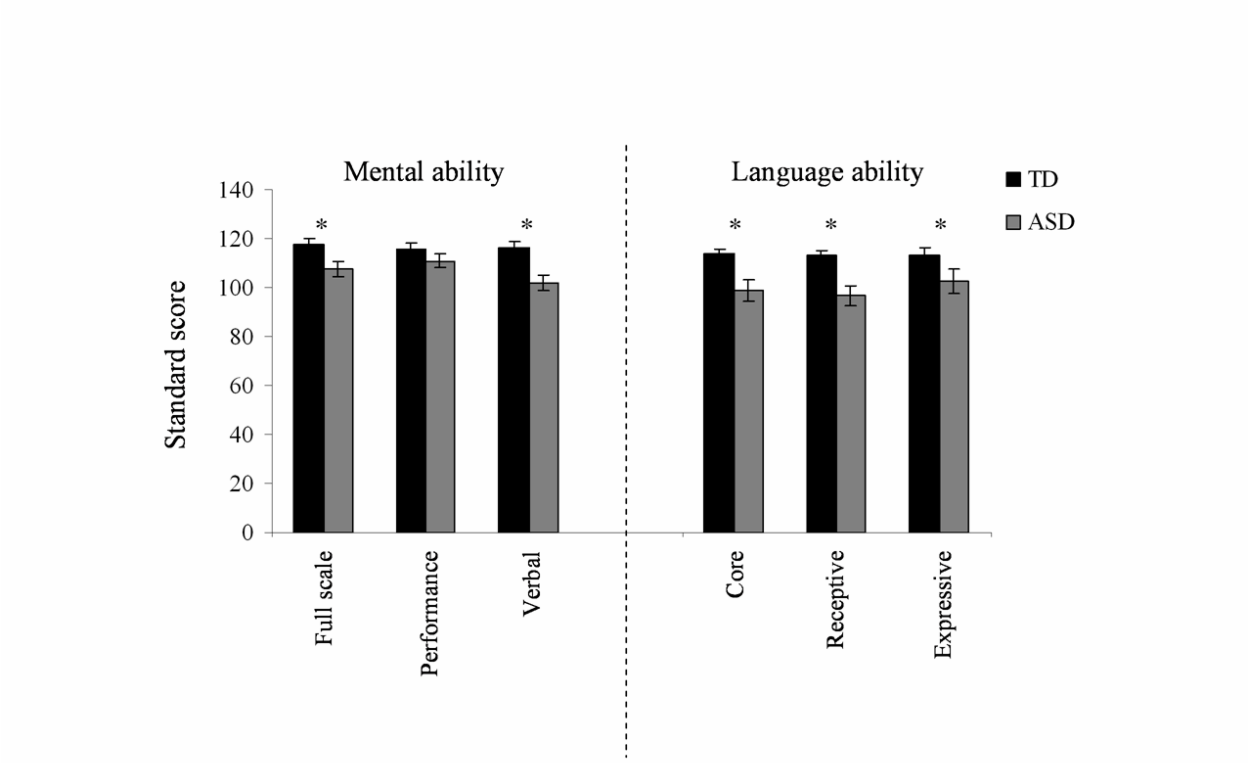

Fig. 8. Mental (left) and language ability (right) means (standard errors) for TD and ASD groups. Children with ASD demonstrated poorer mental and language abilities, although their mental ability level was within normal limits.

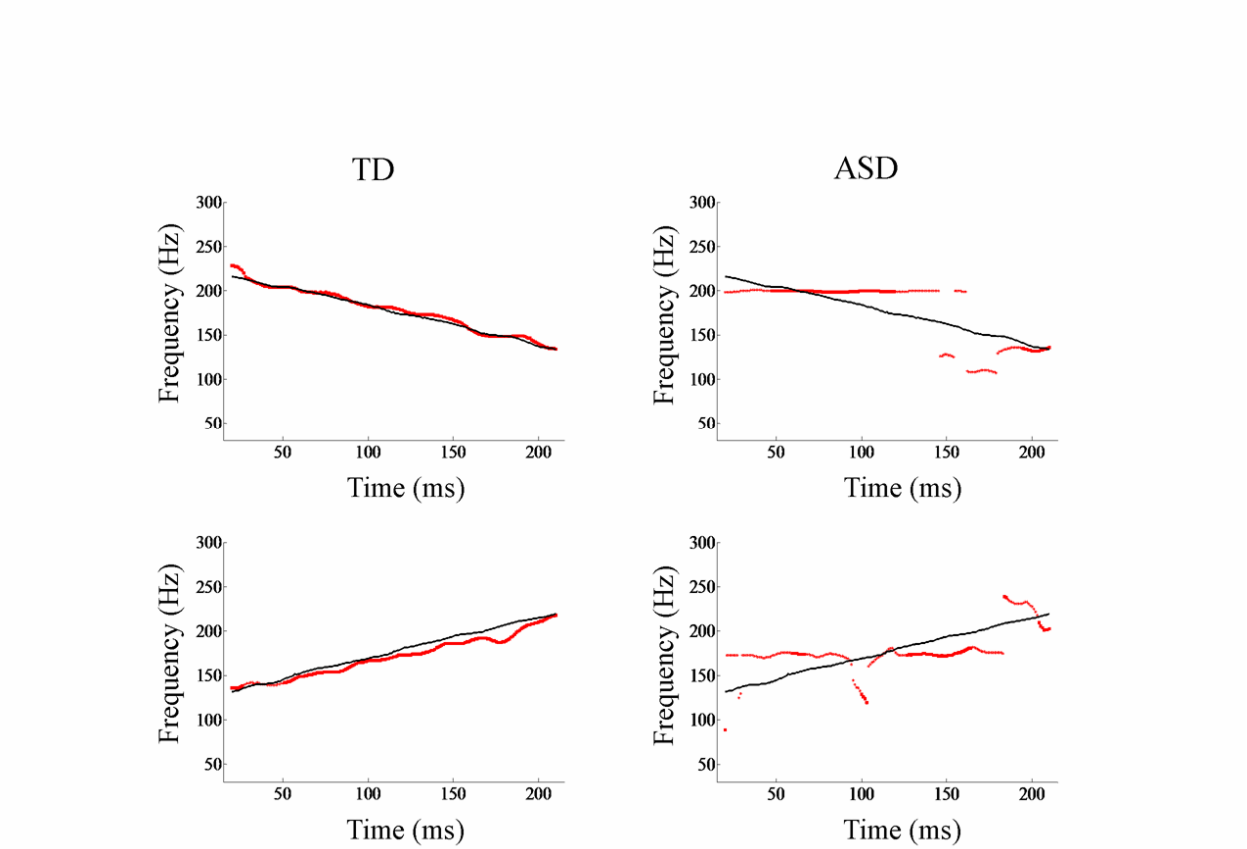

Fig. 9. Representative pitch-tracking contours extracted from brainstem responses of TD (left) and ASD (right) individuals. The fundamental frequency contour of the response (red) is plotted against the contour of the stimulus (black). Shown here are data from both the descending (top) and ascending (bottom) /ya/ stimuli. Pitch tracking is more precise in the typically-developing system. Frequency (Hz) is plotted along the y-axis. The x-axis shows the time corresponding to the midpoint of each 40-ms time bin analyzed.

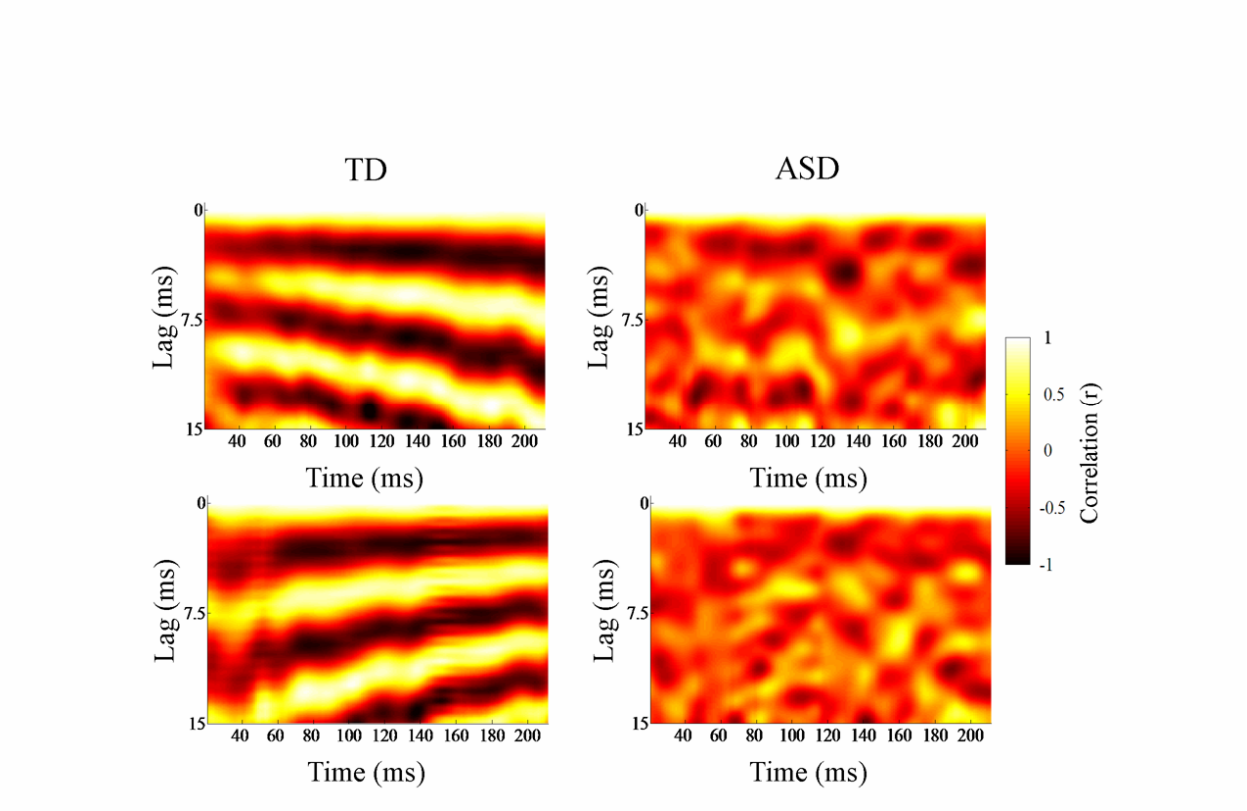

Fig. 10. Autocorrelograms of individual TD (left) and ASD (right) brainstem responses to descending (top) and ascending (bottom) /ya/ stimuli. Running autocorrelations quantify the degree of neural phase locking over time. The autocorrelograms (lag versus time) act a means of visualizing periodicity variation over the course of the response. The time indicated on the x-axis refers to the midpoint of each 40-ms time bin analyzed. The y-axis refers to the amount of lag between the signal (each 40-ms time bin) and a time-shifted copy, and the third dimension, Pitch Strength, is plotted using a color continuum from black to white, with brighter colors representing higher correlations, or more robust encoding of the fundamental frequency contour. The TD response indicates more precise phase locking of pitch than the ASD response.

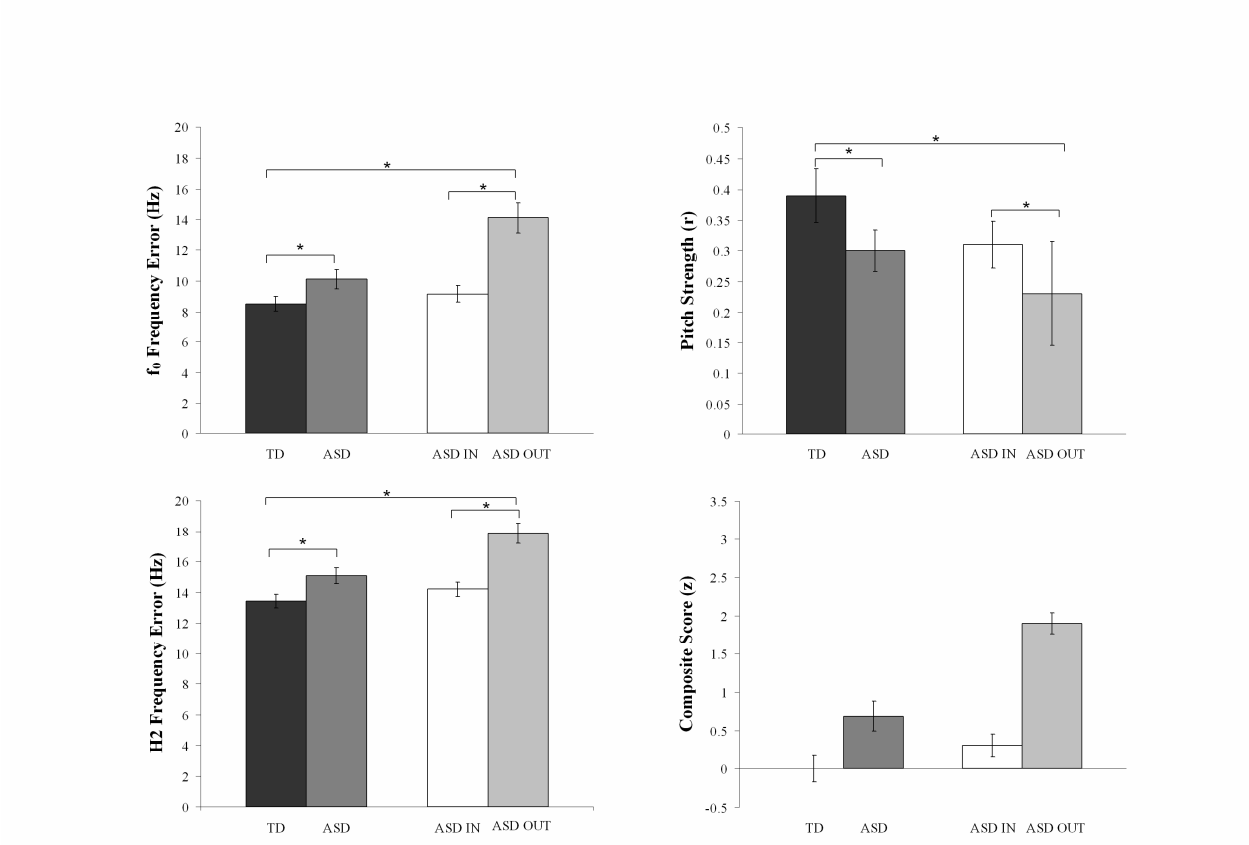

Fig. 11. Group means (standard error) for f0 Frequency Error (Hz), Pitch Strength (autocorrelation r values), H2 Frequency Error (Hz) and Composite Score (z values). Encoding was significantly more precise in TD responses (left, black) as compared to the ASD group as a whole (middle left, dark gray). ASD OUT children (light gray) are those who have pitch tracking composite scores outside of the TD group, while ASD IN children (middle right, white) have scores that are within the normal range. The ASD OUT group (far right, gray) was largely driving the significant group differences, as the ASD IN group demonstrated encoding similar to the TD group.

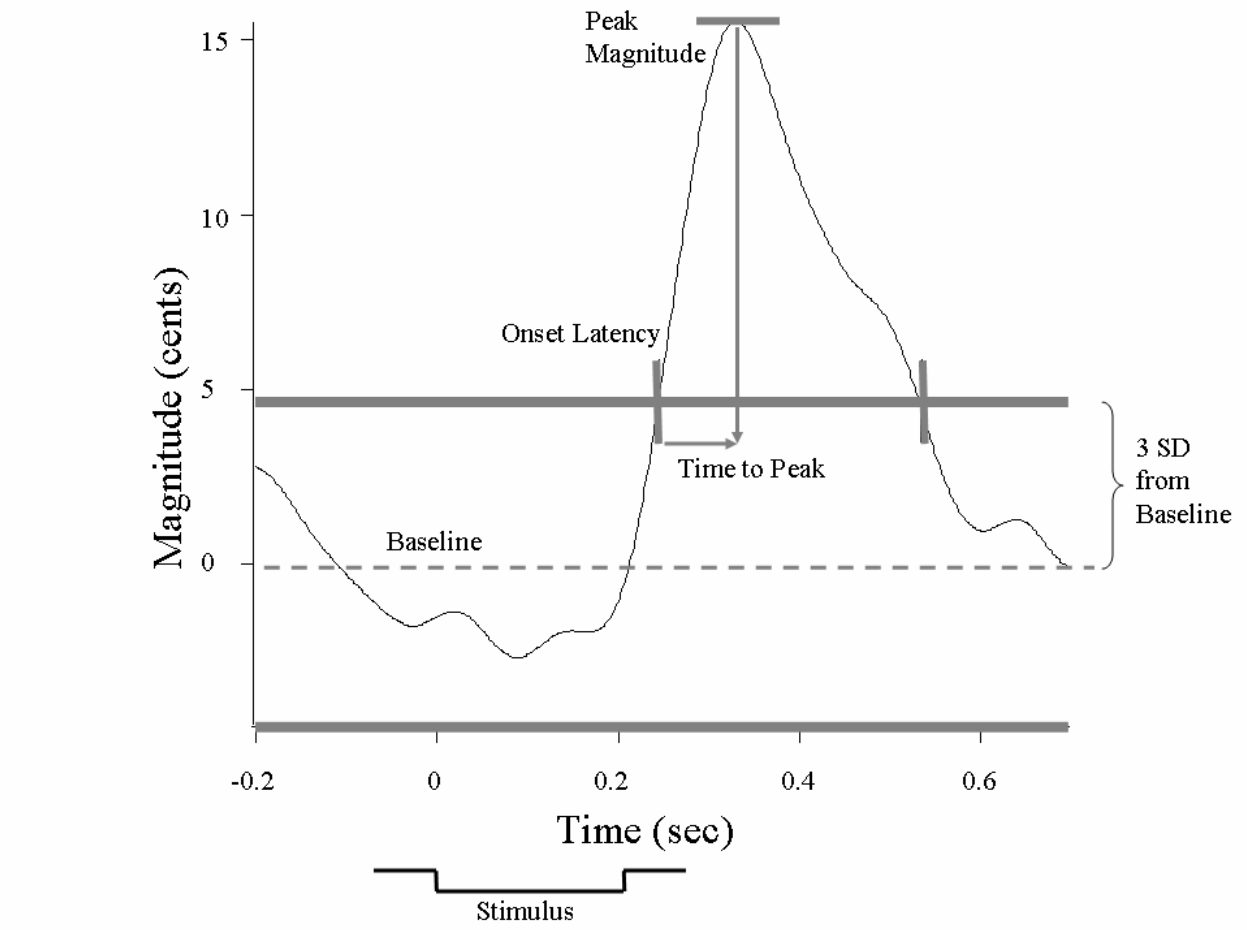

Fig. 12. Exemplar vocal response to pitch perturbation during auditory feedback. The stimulus (indicated underneath the x-axis) begins at time 0 and lasts 200 ms. To evaluate the pitch-shift reflex, first the baseline mean frequency (cents) and three standard deviations (SD) from the mean as threshold were measured. From these benchmarks, the pitch-shift reflex can be evaluated. The time when the frequency exceeds threshold is the "onset latency" of the response; the maximum peak of the response is the peak "magnitude"; the difference between the latency at which the peak magnitude is achieved and the onset latency is the "time to peak"; and the time when the frequency falls below the three SD threshold again represents the end of the

response reflex. The important response features are demarcated in gray color. Note that the y-axis has been de-meaned to 0 cents.

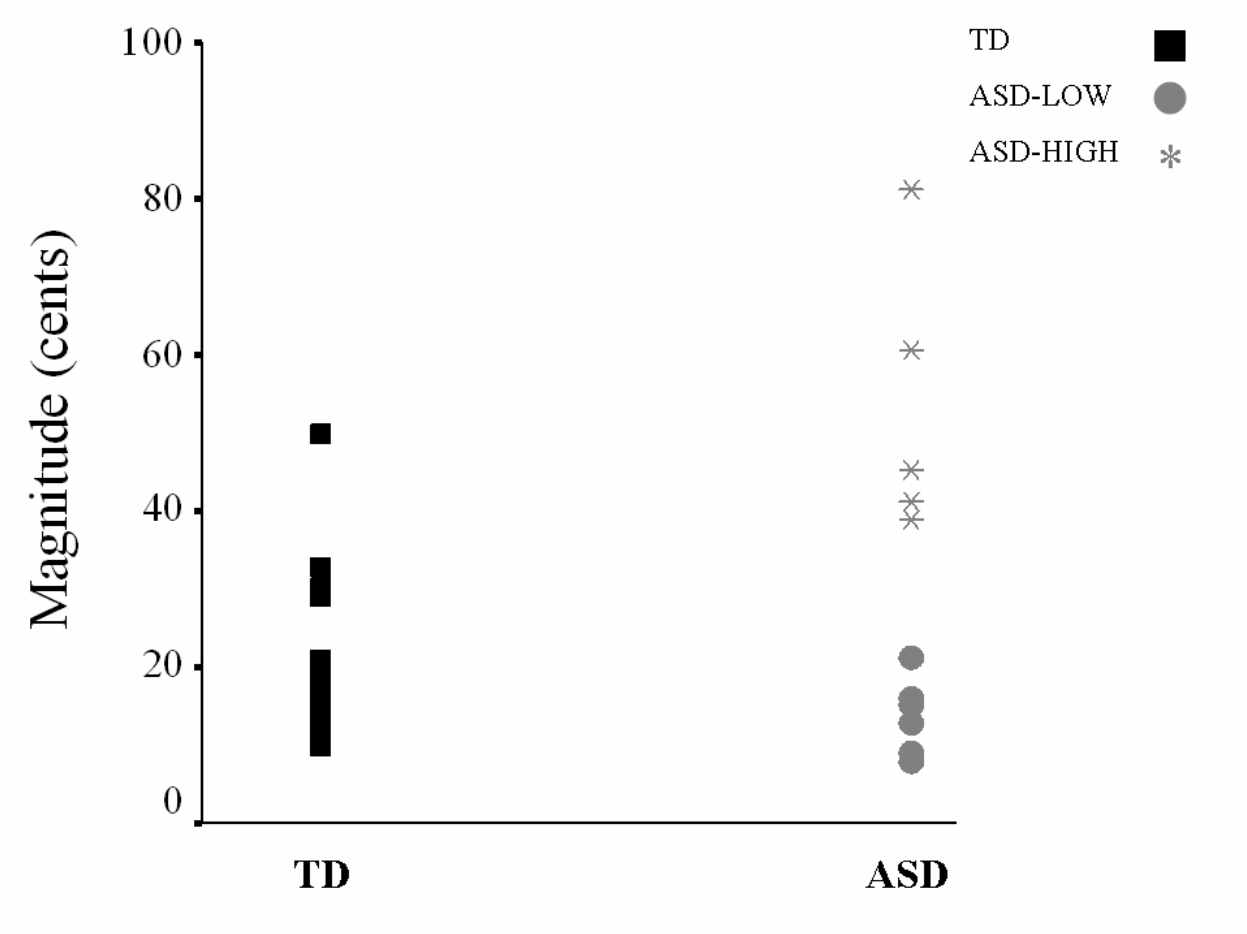

Fig. 13. Dot plot of compensatory response magnitudes (cents) of TD children (left; black squares) and children with ASD (right; ASD-LOW: gray circles and ASD-HIGH: asterisks). Children whose magnitudes were within ±1.65 SD of the typical range comprise the ASD-LOW group, whereas children whose magnitudes exceeded +1.65 SD comprise the ASD-HIGH group.

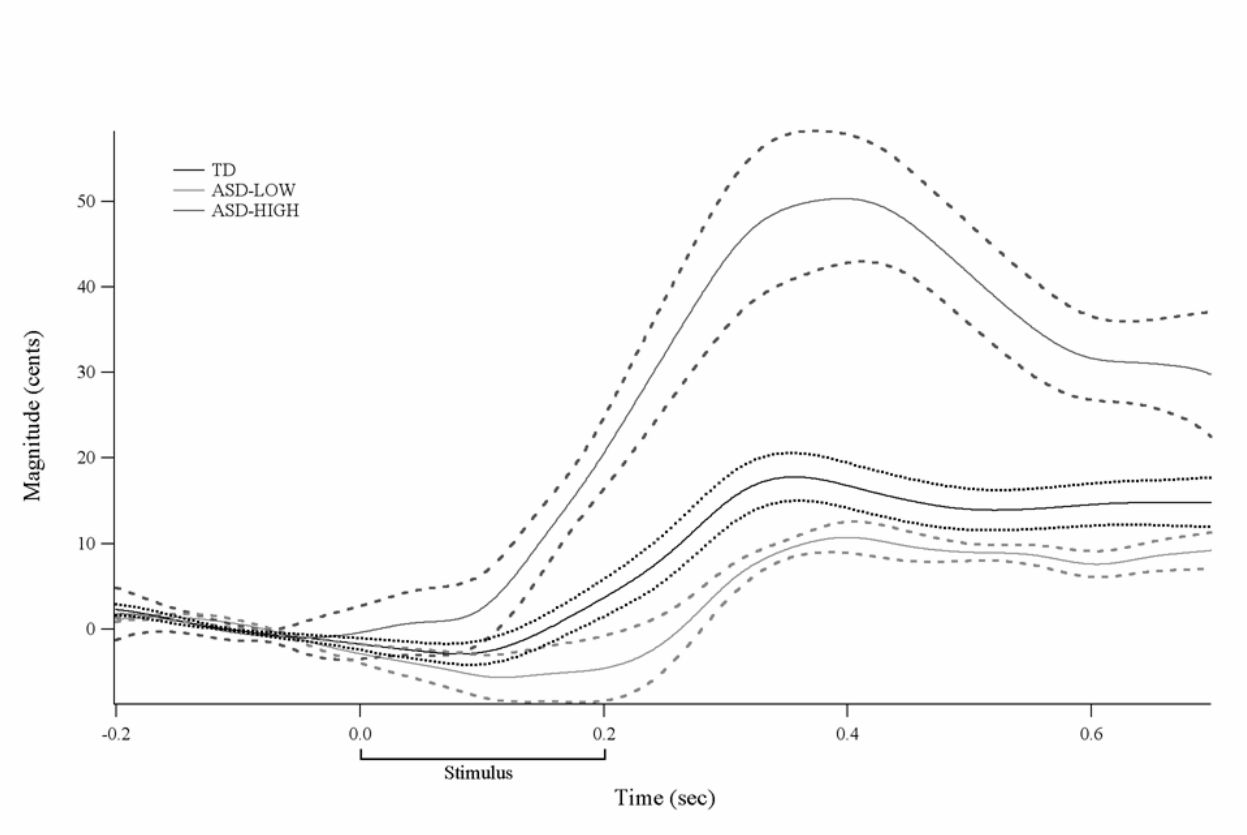

Fig. 14. Grand average TD (black), ASD-LOW (light gray) and ASD-HIGH (dark gray) response magnitude curves and standard error (TD: dotted lines; ASD: dashed lines). Whereas the TD response to pitch perturbation is approximately 20 cents, the ASD-LOW group response is significantly smaller, and the ASD-HIGH group response is significantly larger than the TD group.

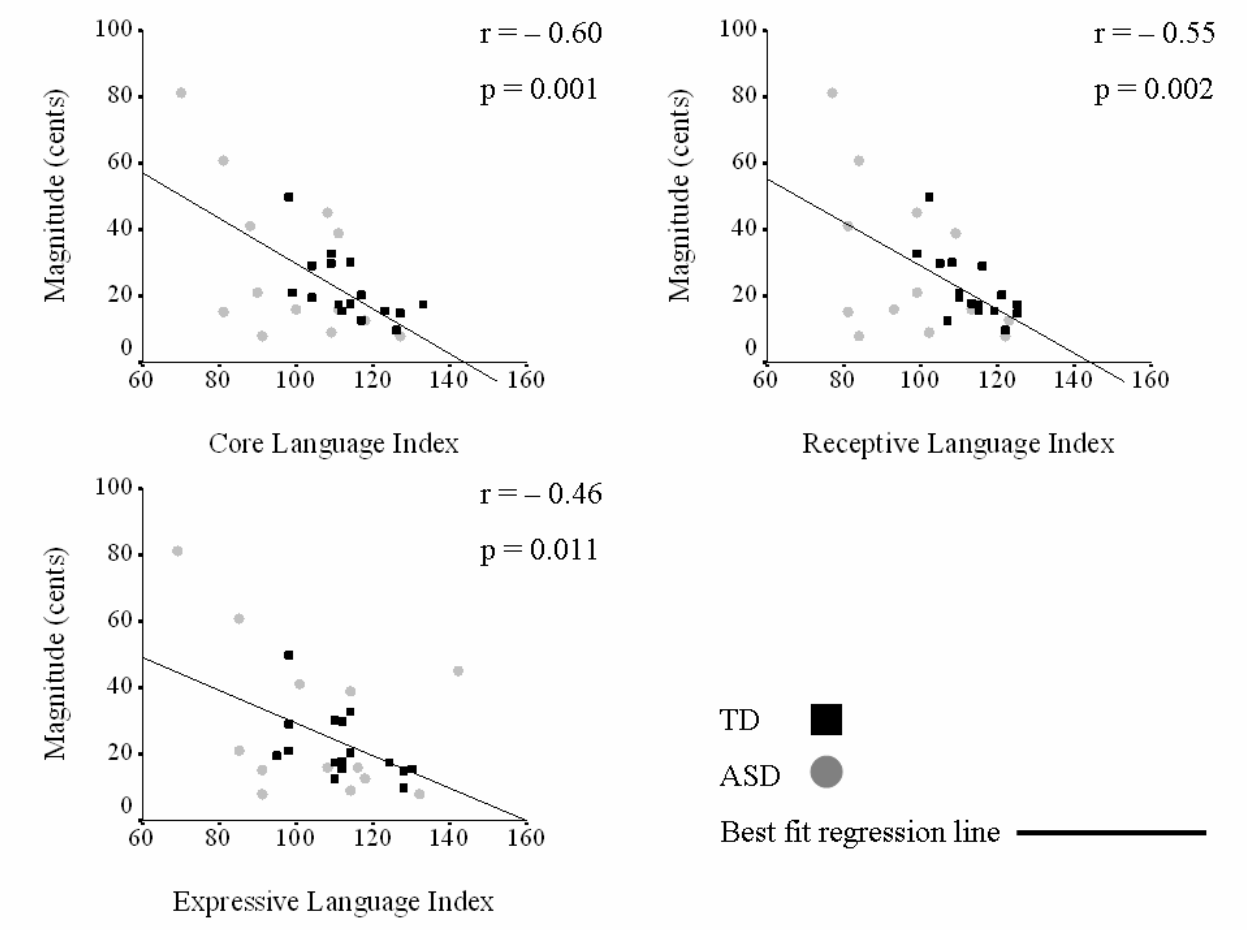

Fig. 15. Relationships between magnitude and language ability. Statistically significant Pearson's correlations ($r \geq 0.32$ and $p \leq 0.05$) were found between magnitude (cents) and core (top left), receptive (top right) and expressive (bottom left) language abilities. Smaller response magnitudes were related to better language scores as measured by the CELF. Magnitudes and behavioral scores of individual subjects (TD, black squares; ASD, gray circles) are plotted, as well as the best fit regression line for the entire sample. Correlation r- and p-values are reported next to each plot.

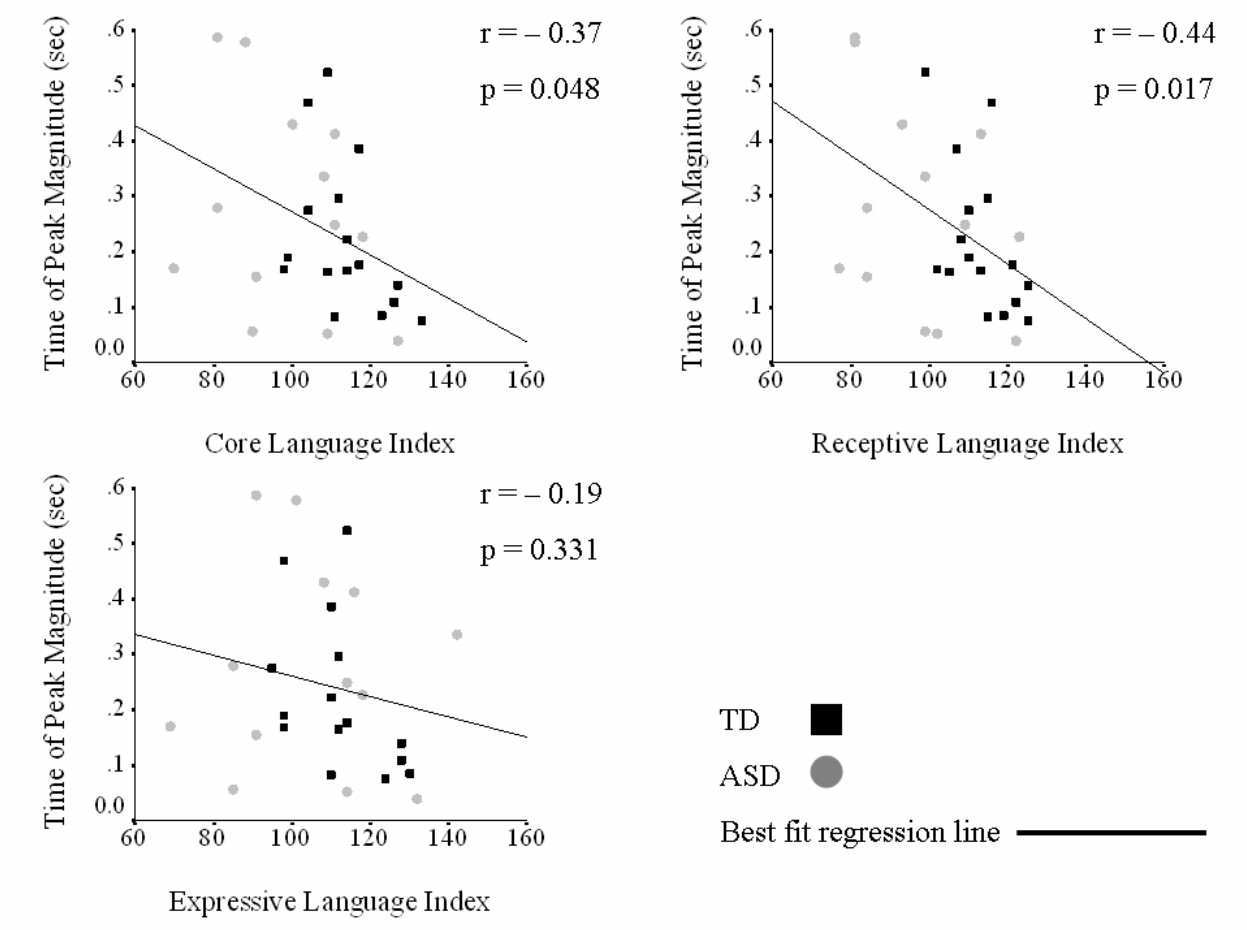

Fig. 16. Relationship between time of peak magnitude and language ability. Statistically significant Pearson's correlations ($r \geq 0.32$ and $p \leq 0.05$) existed between the time of peak magnitude (sec) core (top left) and receptive (top right) language abilities, such that a shorter time to peak was related to better core and receptive language abilities. There was no relationship between time of peak magnitude and expressive language ability (bottom left). Time of peak magnitude and behavioral scores of individual subjects (TD, black squares; ASD, gray circles) are plotted, as well as the best fit regression line for the entire sample. Correlation r- and p-values are reported next to each plot.

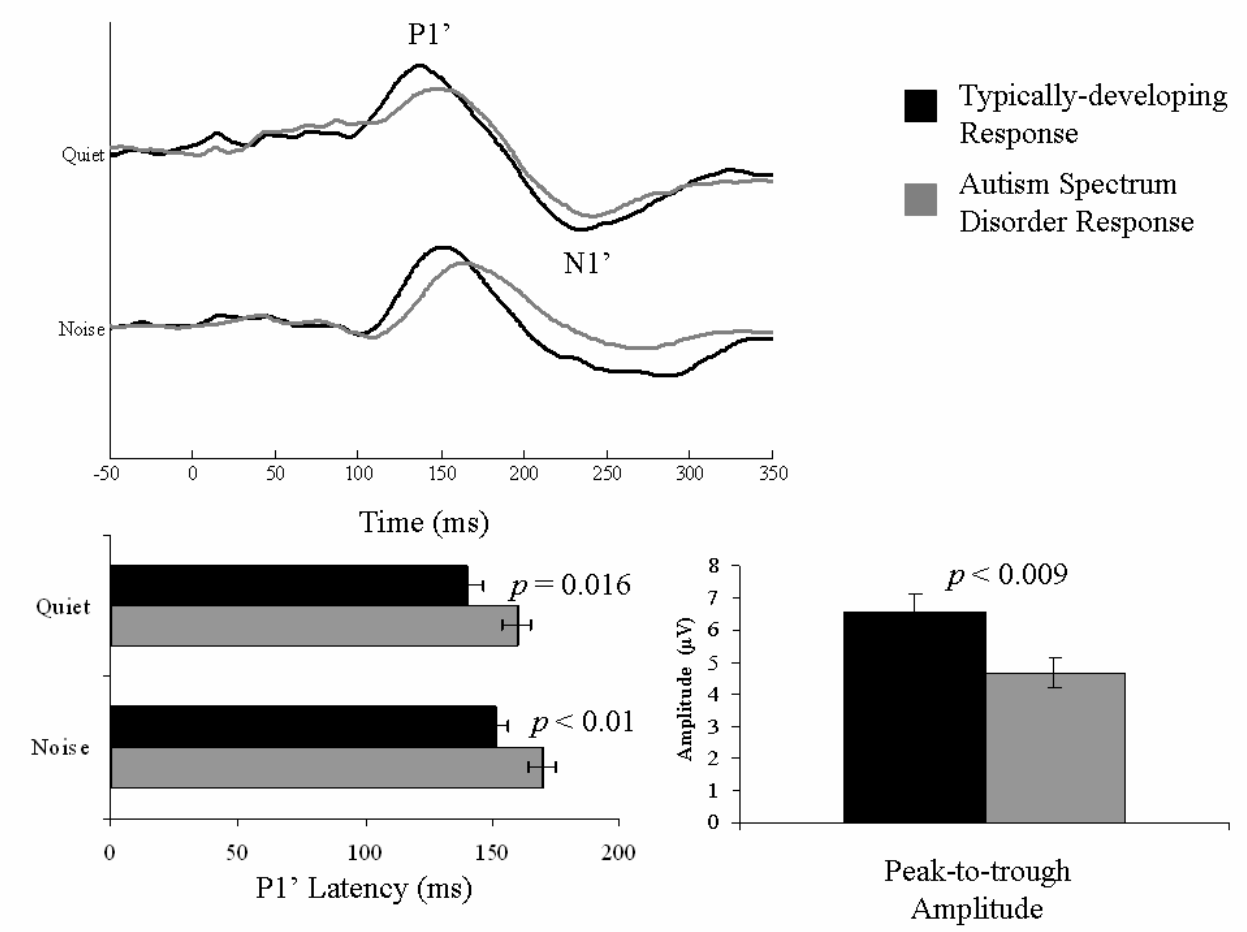

Fig. 17. Grand average cortical responses (quiet (top); background noise (bottom)) of typically-developing (TD) children (black lines) and children with ASD (gray lines). Children with ASD demonstrated significant delays in P1' latency in both the quiet and background noise conditions ($p \leq 0.017$), as well as significant reductions in peak-to-trough amplitudes in the quiet condition.

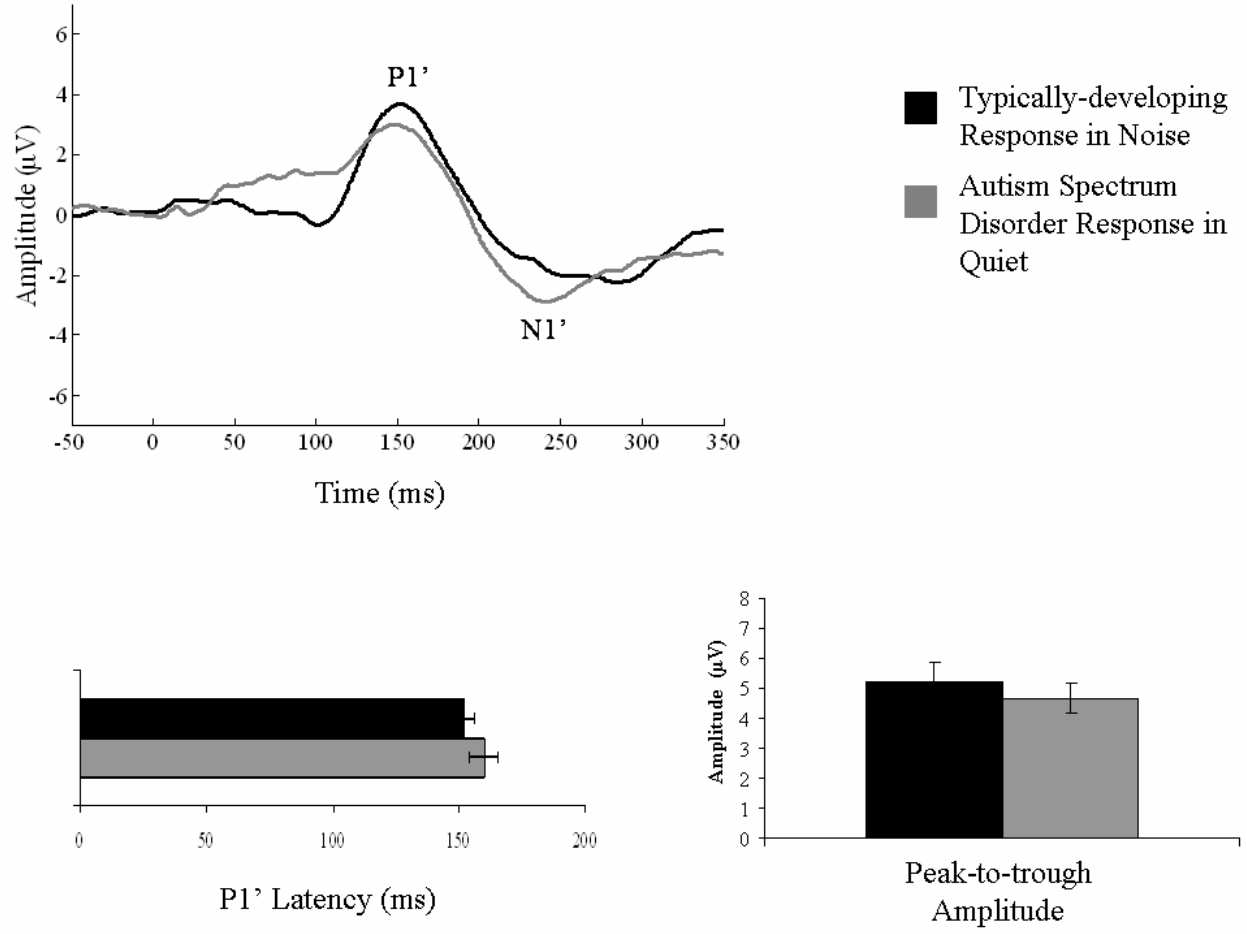

Fig. 18. Comparison of the typical response in background noise to the ASD response in quiet. Responses to stimuli in background noise in the TD group showed no significant differences from the ASD response in quiet. These results suggest that the children with ASD encode speech in quiet similarly to the manner in which TD children encode speech in background noise, giving them a disadvantage for speech perception.

REFERENCES

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends Cogn Sci, 8*(10), 457-464.

Akhoun, I., Gallégo, S., Moulin, A., Ménard, M., Veuillet, E., Berger-Vachon, C., et al. (in press). The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme /ba/ in normal-hearing adults. *Clin Neurophysiol*.

Alcantara, J. I., Weisblatt, E. J., Moore, B. C., & Bolton, P. F. (2004). Speech-in-noise perception in high-functioning individuals with autism or Asperger's syndrome. *J Child Psychol Psychiatry, 45*(6), 1107-1114.

Banai, K., Nicol, T., Zecker, S. G., & Kraus, N. (2005). Brainstem timing: implications for cortical processing and literacy. *J Neurosci, 25*(43), 9850-9857.

Baron-Cohen, S. (2006). The hyper-systemizing, assortative mating theory of autism. *Prog Neuropsychopharmacol Biol Psychiatry, 30*(5), 865-872.

Baron-Cohen, S., Knickmeyer, R. C., & Belmonte, M. K. (2005). Sex Differences in the Brain: Implications for Explaining Autism. *Science, 310*, 819-823.

Bartlett, C. W., Flax, J. F., Logue, M. W., Smith, B. J., Vieland, V. J., Tallal, P., et al. (2004). Examination of potential overlap in autism and language loci on chromosomes 2, 7, and 13 in two independent samples ascertained for specific language impairment. *Hum Hered, 57*(1), 10-20.

Bauer, J. J., & Larson, C. R. (2003). Audio-vocal responses to repetitive pitch-shift stimulation during a sustained vocalization: improvements in methodology for the pitch-shifting technique. *J Acoust Soc Am, 114*(2), 1048-1054.

Baum, S. R., & Pell, M. D. (1999). The neural bases of prosody: Insights from lesion studies and neuroimaging. *Aphasiology, 13*(8), 581-608.

Benasich, A. A., & Tallal, P. (2002). Infant discrimination of rapid auditory cues predicts later language impairment. *Behav Brain Res, 136*(1), 31-49.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., et al. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cereb Cortex, 10*(5), 512-528.

Boddaert, N., Belin, P., Chabane, N., Poline, J. B., Barthelemy, C., Mouren-Simeoni, M. C., et al. (2003). Perception of complex sounds: abnormal pattern of cortical activation in autism. *Am J Psychiatry, 160*(11), 2057-2060.

Boddaert, N., Chabane, N., Belin, P., Bourgeois, M., Royer, V., Barthelemy, C., et al. (2004). Perception of complex sounds in autism: abnormal auditory cortical processing in children. *Am J Psychiatry, 161*(11), 2117-2120.

Boddaert, N., Chabane, N., Gervais, H., Good, C. D., Bourgeois, M., Plumet, M. H., et al. (2004). Superior temporal sulcus anatomical abnormalities in childhood autism: a voxel-based morphometry MRI study. *Neuroimage, 23*(1), 364-369.

Boersma, P., & Weenink, D. (2004). PRAAT: doing phonetics by computer.

Boston, J. R., & Moller, A. R. (1985). Brainstem auditory-evoked potentials. *Crit Rev Biomed Eng, 13*(2), 97-123.

Boucher, J. (2003). Language development in autism. *Int J Pediatr Otorhinolaryngol, 67 Suppl 1*, S159-163.

Boylan, C. B., Blue, M. E., & Hohmann, C. F. (2007). Modeling early cortical serotonerigc deficits in autism. *Behav Brain Res, 176*, 94-108.

Bradlow, A. R., Kraus, N., & Hayes, E. (2003). Speaking clearly for children with learning disabilities: sentence perception in noise. *J Speech Lang Hear Res, 46*(1), 80-97.

Brandt, J., & Rosen, J. J. (1980). Auditory phonemic perception in dyslexia: categorical identification and discrimination of stop consonants. *Brain Lang, 9*(2), 324-337.

Bruneau, N., Bonnet-Brilhault, F., Gomot, M., Adrien, J. L., & Barthelemy, C. (2003). Cortical auditory processing and communication in children with autism: electrophysiological/behavioral relations. *Int J Psychophysiol, 51*(1), 17-25.

Bruneau, N., Roux, S., Adrien, J. L., & Barthelemy, C. (1999). Auditory associative cortex dysfunction in children with autism: evidence from late auditory evoked potentials (N1 wave-T complex). *Clin Neurophysiol, 110*(11), 1927-1934.

Buchwald, J. S., Guthrie, D., Schwafel, J., Erwin, R. J., & Van Lancker, D. (1994). Influence of language structure on brain-behavior development. *Brain Lang, 46*(4), 607-619.

Buchwald, J. S., & Huang, C. (1975). Far-field acoustic response: origins in the cat. *Science, 189*(4200), 382-384.

Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. *J Acoust Soc Am, 103*(6), 3153-3161.

Callan, D. E., Kent, R. D., Guenther, F. H., & Vorperian, H. K. (2000). An auditory-feedback-based neural network model of speech production that is robust to developmental changes in the size and shape of the articulatory system. *J Speech Lang Hear Res, 43*(3), 721-736.

Campisi, P., Low, A., Papsin, B., Mount, R., Cohen-Kerem, R., & Harrison, R. (2005). Acoustic analysis of the voice in pediatric cochlear implant recipients: a longitudinal study. *Laryngoscope, 115*(6), 1046-1050.

Center for Disease Control Prevention. (2007). *Prevalence of autism spectrum disorders--autism and developmental disabilities monitoring network, 14 sites, United States, 2002.*

Ceponiene, R., Lepisto, T., Shestakova, A., Vanhala, R., Alku, P., Naatanen, R., et al. (2003). Speech-sound-selective auditory impairment in children with autism: they can perceive but do not attend. *Proc Natl Acad Sci U S A, 100*(9), 5567-5572.

Ceponiene, R., Shestakova, A., Balan, P., Alku, P., Yiaguchi, K., & Naatanen, R. (2001). Children's auditory event-related potentials index sound complexity and "speechness". *Int J Neurosci, 109*(3-4), 245-260.

Chang, E. F., & Merzenich, M. M. (2003). Environmental noise retards auditory cortical development. *Science, 300*(5618), 498-502.

Chen, S. H., Liu, H., Xu, Y., & Larson, C. R. (2007). Voice F0 responses to pitch-shifted voice feedback during English speech. *J Acoust Soc Am, 121*(2), 1157-1163.

Courchesne, E., Courchesne, R. Y., Hicks, G., & Lincoln, A. J. (1985). Functioning of the brain-stem auditory pathway in non-retarded autistic individuals. *Electroencephalogr Clin Neurophysiol, 61*(6), 491-501.

Courchesne, E., & Pierce, K. (2005). Why the frontal cortex in autism might be talking only to itself: local over-connectivity but long-distance disconnection. *Curr Opin Neurobiol, 15*(2), 225-230.

Cunningham, J., Nicol, T., Zecker, S., & Kraus, N. (2000). Speech-evoked neurophysiologic responses in children with learning problems: development and behavioral correlates of perception. *Ear Hear, 21*(6), 554-568.

Cunningham, J., Nicol, T., Zecker, S. G., Bradlow, A., & Kraus, N. (2001). Neurobiologic responses to speech in noise in children with learning problems: deficits and strategies for improvement. *Clin Neurophysiol, 112*(5), 758-767.

Dawson, G., Osterling, J., Meltzoff, A. N., & Kuhl, P. (2000). Case Study of the Development of an Infant with Autism from Birth to Two Years of Age. *Journal of Applied Developmental Psychology, 21*(3), 299-313.

de, A. F., & Hamilton, C. (2008). Emulation and mimicry for social interaction: A theoretical approach to imitation in autism. *The Quarterly Journal of Experimental Psychology, 61*(1), 101-115.

De Giacomo, A., & Fombonne, E. (1998). Parental recognition of developmental abnormalities in autism. *Eur Child Adolesc Psychiatry, 7*(3), 131-136.

de Villers-Sidani, E., Chang, E. F., Bao, S., & Merzenich, M. M. (2007). Critical period window for spectral tuning defined in the primary auditory cortex (A1) in the rat. *J Neurosci, 27*(1), 180-189.

Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: common themes and mechanisms. *Annu Rev Neurosci, 22*, 567-631.

Edeline, J. M., & Weinberger, N. M. (1991a). Subcortical adaptive filtering in the auditory system: associative receptive field plasticity in the dorsal medial geniculate body. *Behav Neurosci, 105*(1), 154-175.

Edeline, J. M., & Weinberger, N. M. (1991b). Thalamic short-term plasticity in the auditory system: associative returning of receptive fields in the ventral medial geniculate body. *Behav Neurosci, 105*(5), 618-639.

Edeline, J. M., & Weinberger, N. M. (1992). Associative retuning in the thalamic source of input to the amygdala and auditory cortex: receptive field plasticity in the medial division of the medial geniculate body. *Behav Neurosci, 106*(1), 81-105.

Ejiri, K. (1998). Relationship between rhythmic behavior and canonical babbling in infant vocal development. *Phonetica, 55*(4), 226-237.

Eliades, S. J., & Wang, X. (2003). Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations. *J Neurophysiol, 89*(4), 2194-2207.

Erwin, R., Van Lancker, D., Guthrie, D., Schwafel, J., Tanguay, P., & Buchwald, J. S. (1991). P3 responses to prosodic stimuli in adult autistic subjects. *Electroencephalogr Clin Neurophysiol, 80*(6), 561-571.

Fant, G. (1960). *Acoustic Theory of Speech Production*. The Hague: Mouton.

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. *Clin Neurophysiol, 114*(9), 1671-1680.

Filipek, P. A., Accardo, P. J., Ashwal, S., Baranek, G. T., Cook, E. H., Jr., Dawson, G., et al. (2000). Practice parameter: screening and diagnosis of autism: report of the Quality

Standards Subcommittee of the American Academy of Neurology and the Child Neurology Society. *Neurology, 55*(4), 468-479.

Fine, I., Finney, E. M., Boynton, G. M., & Dobkins, K. R. (2005). Comparing the effects of auditory deprivation and sign language within the auditory and visual cortex. *J Cogn Neurosci, 17*(10), 1621-1637.

Flagg, E. J., Cardy, J. E., Roberts, W., & Roberts, T. P. (2005). Language lateralization development in children with autism: Insights from the late field magnetoencephalogram. *Neurosci Lett, 386*(2), 82-87.

Freitag, C. M. (2007). The genetics of autistic disorders and its clinical relevance: a review of the literature. *Mol Psychiatry, 12*(1), 2-22.

Frick, R. W. (1985). Communicating Emotion: The Role of Prosodic Features. *Psychological Bulletin, 97*(3), 412-429.

Gabriele, M. L., Brunso-Bechtold, J. K., & Henkel, C. K. (2000). Plasticity in the development of afferent patterns in the inferior colliculus of the rat after unilateral cochlear ablation. *J Neurosci, 20*(18), 6939-6949.

Gage, N. M., Siegel, B., Callen, M., & Roberts, T. P. (2003). Cortical sound processing in children with autism disorder: an MEG investigation. *Neuroreport, 14*(16), 2047-2051.

Gage, N. M., Siegel, B., & Roberts, T. P. (2003). Cortical auditory system maturational abnormalities in children with autism disorder: an MEG investigation. *Brain Res Dev Brain Res, 144*(2), 201-209.

Galbraith, G. C. (1994). Two-channel brain-stem frequency-following responses to pure tone and missing fundamental stimuli. *Electroencephalogr Clin Neurophysiol, 92*(4), 321-330.

Galbraith, G. C., Amaya, E. M., de Rivera, J. M., Donan, N. M., Duong, M. T., Hsu, J. N., et al. (2004). Brain stem evoked response to forward and reversed speech in humans. *Neuroreport, 15*(13), 2057-2060.

Galbraith, G. C., Gutterson, R. P., Levy, D. S., Mussey, J. L., Sabatasso, F. A., & Wasserman, R. I. (2004). Correlated brain stem and cortical evoked responses to auditory tone change. *Neuroreport, 15*(17), 2613-2616.

Galbraith, G. C., Olfman, D. M., & Huffman, T. M. (2003). Selective attention affects human brain stem frequency-following response. *Neuroreport, 14*(5), 735-738.

Galbraith, G. C., Philippart, M., & Stephen, L. M. (1996). Brainstem frequency-following responses in Rett syndrome. *Pediatr Neurol, 15*(1), 26-31.

Gentili, R., Papaxanthis, C., & Pozzo, T. (2006). Improvement and generalization of arm motor performance through motor imagery practice. *Neuroscience, 137*(3), 761-772.

Gernsbacher, M. A. (2004). Language is More than Speech: A Case Study. *The Journal of Developmental and Learning Disorders, 8*, 79-96.

Gervais, H., Belin, P., Boddaert, N., Leboyer, M., Coez, A., Sfaello, I., et al. (2004). Abnormal cortical voice processing in autism. *Nat Neurosci, 7*(8), 801-802.

Gomes, E., Rotta, N. T., Pedroso, F. S., Sleifer, P., & Danesi, M. C. (2004). Auditory hypersensitivity in children and teenagers with autistic spectrum disorder. *Arq Neuropsiquiatr, 62*(3B), 797-801.

Gorga, M. P., Abbas, P. J., & Worthington, D. W. (1985). Stimulus Calibrations in ABR Measurements. In J. T. Jacobson (Ed.), *The Auditory Brainstem Response* (pp. 49-62). San Diego: College-Hill Press.

Gravel, J. S., Dunn, M., Lee, W. W., & Ellis, M. A. (2006). Peripheral audition of children on the autistic spectrum. *Ear Hear, 27*(3), 299-312.

Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *J Commun Disord, 39*(5), 350-365.

Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain Lang, 96*(3), 280-301.

Guenther, F. H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychol Rev, 105*(4), 611-633.

Hain, T. C., Burnett, T. A., Kiran, S., Larson, C. R., Singh, S., & Kenney, M. K. (2000). Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex. *Exp Brain Res, 130*(2), 133-141.

Hall, J. W. (1992). *Handbook of Auditory Evoked Responses*. Needham Heights: Allyn and Bacon.

Hamzavi, J., Deutsch, W., Baumgartner, W. D., Bigenzahn, W., & Gstoettner, W. (2000). Short-term effect of auditory feedback on fundamental frequency after cochlear implantation. *Audiology, 39*(2), 102-105.

Hardan, A. Y., Jou, R. J., Keshavan, M. S., Varma, R., & Minshew, N. J. (2004). Increased frontal cortical folding in autism: a preliminary MRI study. *Psychiatry Res, 131*(3), 263-268.

Hardan, A. Y., Muddasani, S., Vemulapalli, M., Keshavan, M. S., & Minshew, N. J. (2006). An MRI study of increased cortical thickness in autism. *Am J Psychiatry, 163*(7), 1290-1292.

Hashimoto, T., Tayama, M., Miyazaki, M., Murakawa, K., Shimakawa, S., Yoneda, Y., et al. (1993). Brainstem involvement in high functioning autistic children. *Acta Neurol Scand, 88*(2), 123-128.

Hashimoto, T., Tayama, M., Murakawa, K., Yoshimoto, T., Miyazaki, M., Harada, M., et al. (1995). Development of the brainstem and cerebellum in autistic patients. *J Autism Dev Disord, 25*(1), 1-18.

Hayes, E. A., Warrier, C. M., Nicol, T. G., Zecker, S. G., & Kraus, N. (2003). Neural plasticity following auditory training in children with learning problems. *Clin Neurophysiol, 114*(4), 673-684.

Henkel, C. K., Gabriele, M. L., & McHaffie, J. G. (2005). Quantitative assessment of developing afferent patterns in the cat inferior colliculus revealed with calbindin immunohistochemistry and tract tracing methods. *Neurosci, 136*(3), 945-955.

Herbert, M. R., Harris, G. J., Adrien, K. T., Ziegler, D. A., Makris, N., Kennedy, D. N., et al. (2002). Abnormal asymmetry in language association cortex in autism. *Ann Neurol, 52*(5), 588-596.

Herbert, M. R., & Kenet, T. (2007). Brain abnormalities in language disorders and in autism. *Pediatr Clin North Am, 54*(3), 563-583, vii.

Herbert, M. R., Ziegler, D. A., Deutsch, C. K., O'Brien, L. M., Kennedy, D. N., Filipek, P. A., et al. (2005). Brain asymmetries in autism and developmental language disorder: a nested whole-brain analysis. *Brain, 128*(Pt 1), 213-226.

Higgins, M. B., McCleary, E. A., & Schulte, L. (1999). Altered phonatory physiology with short-term deactivation of children's cochlear implants. *Ear Hear, 20*(5), 426-438.

Hobson, R. P. (1986). The autistic child's appraisal of expressions of emotion. *J Child Psychol Psychiatry, 27*(3), 321-342.

Hoelzley, P. D. (1993). Communication potentiating sounds: Developing channels of communication with autistic children through psychobiological responses to novel sound stimuli. *Canadian Journal of Music Therapy, 1*, 54-76.

Hood, L. (1998). *Clinical applications of the auditory brainstem response*. San Diego: Singular Publishing Group, Inc.

Hoormann, J., Falkenstein, M., Hohnsbein, J., & Blanke, L. (1992). The human frequency-following response (FFR): normal variability and relation to the click-evoked brainstem response. *Hear Res, 59*(2), 179-188.

Houde, J. F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: an MEG study. *J Cogn Neurosci, 14*(8), 1125-1138.

Huang, C. M., & Buchwald, J. S. (1979). C comparison of acoustic nerve and cochlear nucleus responses during acoustic habituation. *Brain Res, 173*(3), 562-566.

Huang, C. M., & Buchwald, J. S. (1980). Changes of acoustic nerve and cochlear nucleus evoked potentials due to repetitive stimulation. *Electroencephalogr Clin Neurophysiol, 49*(1-2), 15-22.

Hubbard, K., & Trauner, D. A. (2007). Intonation and emotion in autistic spectrum disorders. *J Psycholinguist Res, 36*(2), 159-173.

Iarocci, G., & McDonald, J. (2006). Sensory integration and the perceptual experience of persons with autism. *J Autism Dev Disord, 36*(1), 77-90.

Iverson, J. M., & Wozniak, R. H. (2007). Variation in vocal-motor development in infant siblings of children with autism. *J Autism Dev Disord, 37*(1), 158-170.

Jacobson, J. T. (Ed.). (1985). *The Auditory Brainstem Response*. San Diego: College-Hill Press.

Jancke, L., Gaab, N., Wustenberg, T., Scheich, H., & Heinze, H. J. (2001). Short-term functional plasticity in the human auditory cortex: an fMRI study. *Brain Res Cogn Brain Res, 12*(3), 479-485.

Jansson-Verkasalo, E., Ceponiene, R., Kielinen, M., Suominen, K., Jantti, V., Linna, S. L., et al. (2003). Deficient auditory processing in children with Asperger Syndrome, as indexed by event-related potentials. *Neurosci Lett, 338*(3), 197-200.

Jerger, J., & Hall, J. (1980). Effects of age and sex on auditory brainstem response. *Arch Otolaryngol, 106*(7), 387-391.

Johnson, K. L., Nicol, T. G., & Kraus, N. (2005). Brain stem response to speech: a biological marker of auditory processing. *Ear Hear, 26*(5), 424-434.

Johnson, K. L., Nicol, T. G., Zecker, S. G., & Kraus, N. (2007). Auditory brainstem correlates of perceptual timing deficits. *J Cogn Neurosci, 19*(3), 376-385.

Just, M. A., Cherkassky, V. L., Keller, T. A., & Minshew, N. J. (2004). Cortical activation and synchronization during sentence comprehension in high-functioning autism: evidence of underconnectivity. *Brain, 127*(Pt 8), 1811-1821.

Kasai, K., Hashimoto, O., Kawakubo, Y., Yumoto, M., Kamio, S., Itoh, K., et al. (2005). Delayed automatic detection of change in speech sounds in adults with autism: a magnetoencephalographic study. *Clin Neurophysiol, 116*(7), 1655-1664.

Kellerman, G. A., Fin, J., & Gorman, J. M. (2005). Auditory Abnormalities in Autism: Toward Functional Distinctions Among Findings. *CMS Spectrums, 10*(9), 748-756.

Keuroghlian, A. S., & Knudsen, E. I. (2007). Adaptive auditory plasticity in developing and adult animals. *Prog Neurobiol, 82*(3), 109-121.

Khaladkar, A., Kartik, N., & Vanaja, C. (2005). Speech Burst and Click Evoked ABR [Electronic Version]. *Audiology Online*.

Khalfa, S., Bruneau, N., Roge, B., Georgieff, N., Veuillet, E., Adrien, J. L., et al. (2001). Peripheral auditory asymmetry in infantile autism. *Eur J Neurosci, 13*(3), 628-632.

Khalfa, S., Bruneau, N., Roge, B., Georgieff, N., Veuillet, E., Adrien, J. L., et al. (2004). Increased perception of loudness in autism. *Hear Res, 198*(1-2), 87-92.

King, C., Warrier, C. M., Hayes, E., & Kraus, N. (2002). Deficits in auditory brainstem pathway encoding of speech sounds in children with learning problems. *Neurosci Lett, 319*(2), 111-115.

Klatt, D. (1980). Software for cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America, 67*, 971-975.

Klatte, M., Meis, M., Sukowski, H., & Schick, A. (2007). Effects of irrelevant speech and traffic noise on speech perception and cognitive performance in elementary school children. *Noise Health, 9*(36), 64-74.

Klin, A. (1993). Auditory brainstem responses in autism: brainstem dysfunction or peripheral hearing loss? *J Autism Dev Disord, 23*(1), 15-35.

Knickmeyer, R. C., & Baron-Cohen, S. (2006). Fetal testosterone and sex differences in typical social development and in autism. *J Child Neurol, 21*(10), 825-845.

Korpilahti, P., Jansson-Verkasalo, E., Mattila, M. L., Kuusikko, S., Suominen, K., Rytky, S., et al. (2006). Processing of Affective Speech Prosody is Impaired in Asperger Syndrome. *J Autism Dev Disord*.

Kraus, N., & Banai, K. (2007). Auditory-processing malleability. *Current Directions in Psychological Science, 16*(2), 105-110.

Kraus, N., McGee, T. J., Carrell, T. D., Zecker, S. G., Nicol, T. G., & Koch, D. B. (1996). Auditory neurophysiologic responses and discrimination deficits in children with learning problems. *Science, 273*(5277), 971-973.

Kraus, N., & Nicol, T. (2005). Brainstem origins for cortical 'what' and 'where' pathways in the auditory system. *Trends Neurosci, 28*(4), 176-181.

Krishnan, A., & Parkinson, J. (2000). Human frequency-following response: representation of tonal sweeps. *Audiol Neurootol, 5*(6), 312-321.

Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Brain Res Cogn Brain Res, 25*(1), 161-168.

Krishnan, A., Xu, Y., Gandour, J. T., & Cariani, P. A. (2004). Human frequency-following response: representation of pitch contours in Chinese tones. *Hear Res, 189*(1-2), 1-12.

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. *Dev Sci, 8*(1), F1-F12.

Kujala, T., Alho, K., & Naatanen, R. (2000). Cross-modal reorganization of human cortical functions. *Trends Neurosci, 23*(3), 115-120.

Kujala, T., Lepisto, T., Nieminen-von Wendt, T., Naatanen, P., & Naatanen, R. (2005). Neurophysiological evidence for cortical discrimination impairment of prosody in Asperger syndrome. *Neurosci Lett, 383*(3), 260-265.

Ladefoged, P. (2001). *A Course in Phonetics*. Fort Worth: Harcourt College Publishers.

Lane, H., & Tranel, B. (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech and Hearing Research, 14*, 677-709.

Lane, H., Wozniak, J., Matthies, M., Svirsky, M., Perkell, J., O'Connell, M., et al. (1997). Changes in sound pressure and fundamental frequency contours following changes in hearing status. *J Acoust Soc Am, 101*(4), 2244-2252.

Lane, H. L., Catania, A. C., & Stevens, S. S. (1961). Voice level: autophonic scale, perceived loudness, and effects of sidetone. *J Acoust Soc Amer 33*(2), 160-167.

Larson, C. R., Sun, J., & Hain, T. C. (2007). Effects of simultaneous perturbation of voice pitch and loudness feedback on voice F0 and amplitude control. *J Acoust Soc Am, 121*(5), 2862-2872.

Le Couteur, A., Rutter, M., Lord, C., Rios, P., Robertson, S., Holdgrafer, M., et al. (1989). Autism diagnostic interview: a standardized investigator-based instrument. *J Autism Dev Disord, 19*(3), 363-387.

Leder, S. B., Spitzer, J. B., Milner, P., Flevaris-Phillips, C., Kirchner, J. C., & Richardson, F. (1987). Voice intensity of prospective cochlear implant candidates and normal hearing adult males. *Laryngoscope, 97*(2), 224-227.

Leonardo, A., & Konishi, M. (1999). Decrystallization of adult birdsong by perturbation of auditory feedback. *Nature, 399*(6735), 466-470.

Lepisto, T., Kajander, M., Vanhala, R., Alku, P., Huotilainen, M., Naatanen, R., et al. (2007). The perception of invariant speech features in children with autism. *Biol Psychol*.

Lepisto, T., Kujala, T., Vanhala, R., Alku, P., Huotilainen, M., & Naatanen, R. (2005). The discrimination of and orienting to speech and non-speech sounds in children with autism. *Brain Res, 1066*(1-2), 147-157.

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

Levitin, D. J. (2006). The Music Instinct. In *This Is Your Brain on Music: The Science of Human Obsession* (pp. 240-261). New York: Dutton.

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

Liebenthal, E., Binder, J. R., Piorkowski, R. L., & Remez, R. E. (2003). Short-term reorganization of auditory analysis induced by phonetic experience. *J Cogn Neurosci, 15*(4), 549-558.

Lincoln, A. J., Courchesne, E., Harms, L., & Allen, M. (1995). Sensory modulation of auditory stimuli in children with autism and receptive developmental language disorder: event-related brain potential evidence. *J Autism Dev Disord, 25*(5), 521-539.

Liu, H., & Larson, C. R. (in press). Effects of perturbation magnitude and voice F0 level on the pitch-shift reflex. *JASA, 122*(6).

London, E. (2007). The role of the neurobiologist in redefining the diagnosis of autism. *Brain Pathol, 17*(4), 408-411.

Lord, C., Pickles, A., McLennan, J., Rutter, M., Bregman, J., Folstein, S., et al. (1997). Diagnosing autism: analyses of data from the Autism Diagnostic Interview. *J Autism Dev Disord, 27*(5), 501-517.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Jr., Leventhal, B. L., DiLavore, P. C., et al. (2000). The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Dev Disord, 30*(3), 205-223.

Lord, C., Rutter, M., Goode, S., Heemsbergen, J., Jordan, H., Mawhood, L., et al. (1989). Autism diagnostic observation schedule: a standardized observation of communicative and social behavior. *J Autism Dev Disord, 19*(2), 185-212.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord, 24*(5), 659-685.

Magne, C., Schon, D., & Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: behavioral and electrophysiological approaches. *J Cogn Neurosci, 18*(2), 199-211.

Margoliash, D. (2002). Evaluating theories of bird song learning: implications for future directions. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol, 188*(11-12), 851-866.

Marler, P., & Sherman, V. (1983). Song structure without auditory feedback: emendations of the auditory template hypothesis. *J Neurosci, 3*(3), 517-531.

Martin, B. A., Sigal, A., Kurtzberg, D., & Stapells, D. R. (1997). The effects of decreased audibility produced by high-pass noise masking on cortical event-related potentials to speech sounds/ba/and/da. *J Acoust Soc Am, 101*(3), 1585-1599.

Maziade, M., Merette, C., Cayer, M., Roy, M. A., Szatmari, P., Cote, R., et al. (2000). Prolongation of brainstem auditory-evoked responses in autistic probands and their unaffected relatives. *Arch Gen Psychiatry, 57*(11), 1077-1083.

McCabe, K. A., Smith, V. L., & LePore, M. (2000). Intentionality detection and "mindreading": why does game form matter? *Proc Natl Acad Sci U S A, 97*(8), 4404-4409.

McCann, J., & Peppe, S. (2003). Prosody in autism spectrum disorders: a critical review. *Int J Lang Commun Disord, 38*(4), 325-350.

McCann, J., Peppe, S., Gibbon, F. E., O'Hare, A., & Rutherford, M. (2007). Prosody and its relationship to language in school-aged children with high-functioning autism. *Int J Lang Commun Disord, 42*(6), 682-702.

McClelland, R. J., Eyre, D. G., Watson, D., Calvert, G. J., & Sherrard, E. (1992). Central conduction time in childhood autism. *Br J Psychiatry, 160*, 659-663.

McGovern, C. W., & Sigman, M. (2005). Continuity and change from early childhood to adolescence in autism. *J Child Psychol Psychiatry, 46*(4), 401-408.

Miller, C. A. (2006). Developmental relationships between language and theory of mind. *Am J Speech Lang Pathol, 15*(2), 142-154.

Miller, S. B., & Toca, J. M. (1979). Adapted melodic intonation on therapy: A case study of an experimental language program for an autistic child. *J Clin Psychol, 40*, 201-203.

Minshew, N. J., & Hobson, J. A. (2008). Sensory Sensitivities and Performance on Sensory Perceptual Tasks in High-functioning Individuals with Autism. *J Autism Dev Disord*.

Minshew, N. J., & Williams, D. L. (2007). The new neurobiology of autism: cortex, connectivity, and neuronal organization. *Arch Neurol, 64*(7), 945-950.

Moller, A., & Jannetta, P. (1985). Neural generators of the auditory brainstem response. In J. T. Jacobson (Ed.), *The auditory brainstem response* (pp. 13-31). San Diego: College-Hill Press.

Monini, S., Banci, G., Barbara, M., Argiro, M. T., & Filipo, R. (1997). Clarion cochlear implant: short-term effects on voice parameters. *Am J Otol, 18*(6), 719-725.

Moreno, S., & Besson, M. (2005). Influence of musical training on pitch processing: event-related brain potential studies of adults and children. *Ann N Y Acad Sci, 1060*, 93-97.

Muller, R. A., Behen, M. E., Rothermel, R. D., Chugani, D. C., Muzik, O., Mangner, T. J., et al. (1999). Brain mapping of language and auditory perception in high-functioning autistic adults: a PET study. *J Autism Dev Disord, 29*(1), 19-31.

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc Natl Acad Sci U S A, 104*(40), 15894-15898.

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

Nagarajan, S., Mahncke, H., Salz, T., Tallal, P., Roberts, T., & Merzenich, M. M. (1999). Cortical auditory signal processing in poor readers. *Proc Natl Acad Sci U S A, 96*(11), 6483-6488.

Nakahara, H., Zhang, L. I., & Merzenich, M. M. (2004). Specialization of primary auditory cortex processing by sound exposure in the "critical period". *Proc Natl Acad Sci U S A, 101*(18), 7170-7174.

Natke, U., Donath, T. M., & Kalveram, K. T. (2003). Control of voice fundamental frequency in speaking versus singing. *J Acoust Soc Am, 113*(3), 1587-1593.

O'Neill, M., & Jones, R. S. (1997). Sensory-perceptual abnormalities in autism: a case for more research? *J Autism Dev Disord, 27*(3), 283-293.

Oades, R. D., Walker, M. K., Geffen, L. B., & Stern, L. M. (1988). Event-related potentials in autistic and healthy children on an auditory choice reaction time task. *Int J Psychophysiol, 6*(1), 25-37.

Oberman, L. M., & Ramachandran, V. S. (2007). The simulating social mind: the role of the mirror neuron system and simulation in the social and communicative deficits of autism spectrum disorders. *Psychol Bull, 133*(2), 310-327.

Oram Cardy, J. E., Flagg, E. J., Roberts, W., Brian, J., & Roberts, T. P. (2005). Magnetoencephalography identifies rapid temporal processing deficit in autism and language impairment. *Neuroreport, 16*(4), 329-332.

Patel, A. D., Peretz, I., Tramo, M., & Labreque, R. (1998). Processing prosodic and musical patterns: a neuropsychological investigation. *Brain Lang, 61*(1), 123-144.

Paul, R., Augustyn, A., Klin, A., & Volkmar, F. R. (2005). Perception and production of prosody by speakers with autism spectrum disorders. *J Autism Dev Disord, 35*(2), 205-220.

Paul, R., Shriberg, L. D., McSweeny, J., Cicchetti, D., Klin, A., & Volkmar, F. (2005). Brief report: relations between prosodic performance and communication and socialization ratings

in high functioning speakers with autism spectrum disorders. *J Autism Dev Disord, 35*(6), 861-869.

Peppe, S., McCann, J., Gibbon, F., O'Hare, A., & Rutherford, M. (2007). Receptive and expressive prosodic ability in children with high-functioning autism. *J Speech Lang Hear Res, 50*(4), 1015-1028.

Perkell, J., Lane, H., Svirsky, M., & Webster, J. (1992). Speech of cochlear implant patients: a longitudinal study of vowel production. *J Acoust Soc Am, 91*(5), 2961-2978.

Picard, M., & Bradley, J. S. (2001). Revisiting speech interference in classrooms. *Audiology, 40*(5), 221-244.

Plaisted, K., Saksida, L., Alcantara, J., & Weisblatt, E. (2003). Towards an understanding of the mechanisms of weak central coherence effects: experiments in visual configural learning and auditory perception. *Philos Trans R Soc Lond B Biol Sci, 358*(1430), 375-386.

Prather, J. F., Peters, S., Nowicki, S., & Mooney, R. (2008). Precise auditory-vocal mirroring in neurons for learned vocal communication. *Nature, 451*(7176), 305-310.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behav Brain Sci, 4*, 515-526.

Quian Quiroga, R., & Garcia, H. (2003). Single-trial event-related potentials with wavelet denoising. *Clin Neurophysiol, 114*(2), 376-390.

Rapin, I., & Dunn, M. (2003). Update on the language disorders of individuals on the autistic spectrum. *Brain Dev, 25*(3), 166-172.

Recanzone, G. H., Schreiner, C. E., & Merzenich, M. M. (1993). Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *J Neurosci, 13*(1), 87-103.

Rodier, P. M. (2000). The early origins of autism. *Sci Am, 282*(2), 56-63.

Rodier, P. M., Ingram, J. L., Tisdale, B., & Croog, V. J. (1997). Linking etiologies in humans and animal models: studies of autism. *Reprod Toxicol, 11*(2-3), 417-422.

Rojas, D. C., Bawn, S. D., Benkers, T. L., Reite, M. L., & Rogers, S. J. (2002). Smaller left hemisphere planum temporale in adults with autistic disorder. *Neurosci Lett, 328*(3), 237-240.

Rojas, D. C., Camou, S. L., Reite, M. L., & Rogers, S. J. (2005). Planum temporale volume in children and adolescents with autism. *J Autism Dev Disord, 35*(4), 479-486.

Rosenhall, U., Nordin, V., Brantberg, K., & Gillberg, C. (2003). Autism and auditory brain stem responses. *Ear Hear, 24*(3), 206-214.

Rosenhall, U., Nordin, V., Sandstrom, M., Ahlsen, G., & Gillberg, C. (1999). Autism and hearing loss. *J Autism Dev Disord, 29*(5), 349-357.

Rumsey, J. M., Grimes, A. M., Pikus, A. M., Duara, R., & Ismond, D. R. (1984). Auditory brainstem responses in pervasive developmental disorders. *Biol Psychiatry, 19*(10), 1403-1418.

Rupa, V., & Dayal, A. K. (1993). Wave V latency shifts with age and sex in normals and patients with cochlear hearing loss: development of a predictive model. *Br J Audiol, 27*(4), 273-279.

Russo, N., Nicol, T., Musacchia, G., & Kraus, N. (2004). Brainstem responses to speech syllables. *Clinical Neurophysiology, 115*(9), 2021-2030.

Russo, N., Nicol, T., Zecker, S., Hayes, E., & Kraus, N. (2005). Auditory training improves neural timing in the human brainstem. *Behav Brain Res, 156*(1), 95-103.

Russo, N., Skoe, E., Trommer, B., Nicol, T., Zecker, S., Bradlow, A., et al. (in press). Deficient Brainstem Encoding of Pitch in Children with Autism Spectrum Disorders. *Clin Neurophysiol*.

Rutter, M. (1974). The development of infantile autism. *Psychol Med, 4*(2), 147-163.

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.

Sanes, D. H., & Constantine-Paton, M. (1985). The sharpening of frequency tuning curves requires patterned activity during development in the mouse, Mus musculus. *J Neurosci, 5*(5), 1152-1166.

Sanes, D. H., & Friaf, E. (2000). Development and influence of inhibition in the lateral superior olivary nucleus. *Hear Res, 147*(1-2), 46-58.

Sankoh, A. J., Huque, M. F., & Dubey, S. D. (1997). Some comments on frequently used multiple endpoint adjustment methods in clinical trials. *Stat Med, 16*(22), 2529-2542.

Schon, D., Magne, C., & Besson, M. (2004). The music of speech: music training facilitates pitch processing in both music and language. *Psychophysiology, 41*(3), 341-349.

Semel, E., Wiig, E. H., & Secord, W. A. (2003). Clinical Evaluation of Language Fundamentals (4th ed.). San Antonio, TX: Harcourt Assessment, Inc.

Seri, S., Cerquiglini, A., Pisani, F., & Curatolo, P. (1999). Autism in tuberous sclerosis: evoked potential evidence for a deficit in auditory sensory processing. *Clin Neurophysiol, 110*(10), 1825-1830.

Shriberg, L. D., Paul, R., McSweeny, J. L., Klin, A. M., Cohen, D. J., & Volkmar, F. R. (2001). Speech and prosody characteristics of adolescents and adults with high-functioning autism and Asperger syndrome. *J Speech Lang Hear Res, 44*(5), 1097-1115.

Siegal, M., & Blades, M. (2003). Language and auditory processing in autism. *Trends Cogn Sci, 7*(9), 378-380.

Sivasankar, M., Bauer, J. J., Babu, T., & Larson, C. R. (2005). Voice responses to changes in pitch of voice or tone auditory feedback. *J Acoust Soc Am, 117*(2), 850-857.

Smith, J. C., Marsh, J. T., & Brown, W. S. (1975). Far-field recorded frequency-following responses: evidence for the locus of brainstem sources. *Electroencephalogr Clin Neurophysiol, 39*(5), 465-472.

Smith, S. D. (2007). Genes, language development, and language disorders. *Ment Retard Dev Disabil Res Rev, 13*(1), 96-105.

Smotherman, M. S. (2007). Sensory feedback control of mammalian vocalizations. *Behav Brain Res, 182*(2), 315-326.

Song, J., Skoe, E., Wong, P. C. M., & Kraus, N. (2007). *Plasticity in the adult human auditory brainstem following short-term linguistic training.* Paper presented at the Human Brain Mapping Conference, Chicago, IL.

Song, J. H., Banai, K., Russo, N. M., & Kraus, N. (2006). On the relationship between speech- and nonspeech-evoked auditory brainstem responses. *Audiol Neurootol, 11*(4), 233-241.

Starr, A., & Don, M. (1988). Brain potentials evoked by acoustic stimuli. In T. Picton (Ed.), *Human event-related potentials. Handbook of electroencephalography and clinical neurophsyiology* (Vol. 3, pp. 97-158). New York: Elsevier.

Steele, S., Joseph, R. M., & Tager-Flusberg, H. (2003). Brief report: developmental change in theory of mind abilities in children with autism. *J Autism Dev Disord, 33*(4), 461-467.

Stegemöller, E. L., Skoe, E., Nicol, T., Warrier, C. M., & Kraus, N. (in press). Music Training and Vocal Production of Speech and Song. *Music Perception*.

Suga, N., Gao, E., Zhang, Y., Ma, X., & Olsen, J. F. (2000). The corticofugal system for hearing: recent progress. *Proc Natl Acad Sci U S A, 97*(22), 11807-11814.

Svirsky, M. A., Lane, H., Perkell, J. S., & Wozniak, J. (1992). Effects of short-term auditory deprivation on speech production in adult cochlear implant users. *J Acoust Soc Am, 92*(3), 1284-1300.

Tabachnick, B., & Fidell, L. (2007a). Multivariate analysis of variance and covariance. In S. Hartman (Ed.), *Using Multivariate Statistics* (4th ed., pp. 243-310). Boston: Allyn & Bacon.

Tabachnick, B., & Fidell, L. (2007b). Principal Components and Factor Analysis. In S. Hartman (Ed.), *Using Multivariate Statistics* (5th ed., pp. 607-675). Boston: Pearson Education.

Tager-Flusberg, H., & Caronna, E. (2007). Language disorders: autism and other pervasive developmental disorders. *Pediatr Clin North Am, 54*(3), 469-481, vi.

Tanguay, P. E., Edwards, R. M., Buchwald, J., Schwafel, J., & Allen, V. (1982). Auditory brainstem evoked responses in autistic children. *Arch Gen Psychiatry, 39*(2), 174-180.

Teder-Salejarvi, W. A., Pierce, K. L., Courchesne, E., & Hillyard, S. A. (2005). Auditory spatial localization and attention deficits in autistic adults. *Brain Res Cogn Brain Res, 23*(2-3), 221-234.

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

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2003). Perceiving prosody in speech. Effects of music lessons. *Ann N Y Acad Sci, 999*, 530-532.

Titze, I. R. (1994). Fluctuations and Perturbations in Vocal Output. In *Principles of Voice Production*. Englewood Cliffs: Prentice Hall.

Tonnquist-Uhlen, I. (1996). Topography of auditory evoked cortical potentials in children with severe language impairment. *Scand Audiol Suppl, 44*, 1-40.

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2007). Neural mechanisms underlying auditory feedback control of speech. *Neuroimage*.

Tremblay, K., Kraus, N., McGee, T., Ponton, C., & Otis, B. (2001). Central auditory plasticity: changes in the N1-P2 complex after speech-sound training. *Ear Hear, 22*(2), 79-90.

Tremblay, K. L., & Kraus, N. (2002). Auditory training induces asymmetrical changes in cortical neural activity. *J Speech Lang Hear Res, 45*(3), 564-572.

Volkmar, F. R., Lord, C., Bailey, A., Schultz, R. T., & Klin, A. (2004). Autism and pervasive developmental disorders. *J Child Psychol Psychiatry, 45*(1), 135-170.

Volman, S. F., & Khanna, H. (1995). Convergence of untutored song in group-reared zebra finches (Taeniopygia guttata). *J Comp Psychol, 109*(3), 211-221.

Wang, A., Dapretto, M., Hariri, A., Sigman, M., & Bookheimer, S. (2001). Processing Affective and Linguistic Prosody in Autism: An fMRI Study. *NeuroImage, 13*(6), 621.

Warrier, C. M., Johnson, K. L., Hayes, E. A., Nicol, T., & Kraus, N. (2004). Learning impaired children exhibit timing deficits and training-related improvements in auditory cortical responses to speech in noise. *Exp Brain Res, 157*(4), 431-441.

Wells, B., & Peppe, S. (2003). Intonation abilities of children with speech and language impairments. *J Speech Lang Hear Res, 46*(1), 5-20.

Whiting, K. A., Martin, B. A., & Stapells, D. R. (1998). The effects of broadband noise masking on cortical event-related potentials to speech sounds /ba/ and /da. *Ear Hear, 19*(3), 218-231.

Wible, B., Nicol, T., & Kraus, N. (2002). Abnormal neural encoding of repeated speech stimuli in noise in children with learning problems. *Clin Neurophysiol, 113*(4), 485-494.

Wible, B., Nicol, T., & Kraus, N. (2004). Atypical brainstem representation of onset and formant structure of speech sounds in children with language-based learning problems. *Biol Psychiatry, 67*(3), 299-317.

Wible, B., Nicol, T., & Kraus, N. (2005). Correlation between brainstem and cortical auditory processes in normal and language-impaired children. *Brain, 128*(Pt 2), 417-423.

Wickelgren, I. (2005). Neurology. Autistic brains out of synch? *Science, 308*(5730), 1856-1858.

Williams, J. H., Whiten, A., Suddendorf, T., & Perrett, D. I. (2001). Imitation, mirror neurons and autism. *Neurosci Biobehav Rev, 25*(4), 287-295.

Woerner, C., & Overstreet, K. (1999). Wechsler Abbreviated Scale of Intelligence (WASI). San Antonio, TX: The Psychological Corporation.

Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience, 10*(4), 420-422.

Worden, F. G., & Marsh, J. T. (1968). Frequency-following (microphonic-like) neural responses evoked by sound. *Electroencephalogr Clin Neurophysiol, 25*(1), 42-52.

Xu, Y., Krishnan, A., & Gandour, J. T. (2006). Specificity of experience-dependent pitch representation in the brainstem. *Neuroreport, 17*(15), 1601-1605.

Yaguez, L., Nagel, D., Hoffman, H., Canavan, A. G. M., Wist, E., & Homberg, V. (1998). A mental route to motor learning: Improving trajectorial kinematics through imagery training. *Behav Br Res, 90*, 95-106.

Yan, W., & Suga, N. (1996). Corticofugal modulation of time-domain processing of biosonar information in bats. *Science, 273*, 1100-1103.

Yu, X., Sanes, D. H., Aristizabal, O., Wadghiri, Y. Z., & Turnbull, D. H. (2007). Large-scale reorganization of the tonotopic map in mouse auditory midbrain revealed by MRI. *Proc Natl Acad Sci U S A 104*, 12193-12198.

Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: auditory-motor interactions in music perception and production. *Nat Rev Neurosci, 8*(7), 547-558.

Zhang, L. I., Bao, S., & Merzenich, M. M. (2002). Disruption of primary auditory cortex by synchronous auditory inputs during a critical period. *Proc Natl Acad Sci U S A, 99*(4), 2309-2314.

Zhou, X., & Merzenich, M. M. (2007). Intensive training in adults refines A1 representations degraded in an early postnatal critical period. *Proc Natl Acad Sci U S A, 104*(40), 15935-15940.

APPENDIX ITEMS

MUSICAL EXPERIENCE SHAPES HUMAN BRAINSTEM ENCODING OF LINGUISTIC PITCH PATTERNS

Published in *Nature Neuroscience*

Patrick C M Wong[a,b], Erika Skoe[a], Nicole M Russo[a,b,e], Tasha Dees[a] & Nina Kraus[a-d]

[a]The Roxelyn and Richard Pepper Department of Communication Sciences and Disorders [b]Northwestern University Interdepartmental Neuroscience Program [c]Department of Neurobiology and Physiology [d]Department of Otolaryngology, Northwestern University [e]These authors contributed equally to this work.

Abstract

Music and speech are very cognitively demanding auditory phenomena generally attributed to cortical rather than subcortical circuitry. We examined brainstem encoding of linguistic pitch and found that musicians show more robust and faithful encoding compared with nonmusicians. These results not only implicate a common subcortical manifestation for two presumed cortical functions, but also a possible reciprocity of corticofugal speech and music tuning, providing neurophysiological explanations for musicians' higher language-learning ability. Both music and spoken language involve the use of functionally and acoustically complex sound and are generally attributed to the neocortex (Rauschecker 1998; Patel 2003; Wong et al 2004; Wong et al. in press). Less is known about how long-term experience using these complex sounds shapes subcortical circuitry and the context specificity and reciprocity of this tuning (Russo et al 2005).

By measuring the frequency following response (FFR), which presumably originates from the auditory brainstem (inferior colliculus) and encodes the energy of the stimulus fundamental frequency (f0) with high fidelity (Greenberg et al 1987), previous work (Krishnan et al 2005) has found increased linguistic pitch pattern encoding in Mandarin speaking subjects relative to English-speaking subjects. These results reflect Mandarin-speaking subjects' long-term exposure to linguistic pitch patterns, as Mandarin Chinese, a tone language, uses pitch to signal word meaning (for example, /ma/ spoken with high or rising pitch patterns means 'mother' or 'numb', respectively). Moreover, similar to research on short-term perceptual learning (Li, Piech,.Gilbert 2004), these results can be viewed as context specific (that is, linguistic experiences, subserved by the cortex, enhance the encoding of linguistic information at the brainstem). The nonspecificity of this long-term usage effect, though largely unknown, is both theoretically interesting and clinically and educationally relevant. Nonspecificity would suggest that either speech- or music-related experience can tune sensory encoding in the auditory brainstem via the corticofugal pathway. Notably, this tuning, whether speech- or music-induced, would enhance all relevant auditory functions (both speech and music) subserved by the rostral brainstem.

We measured FFR responses to linguistic pitch patterns at the rostral brainstem in ten amateur musicians and ten nonmusicians who had no previous exposure to a tone language (see Supplementary Table 1 online). Musicians (instrumentalists) had at least 6 years of continuous musical training (mean = 10.7 years) starting at or before the age of 12. Nonmusicians had no more than 3 years (mean =1.2 years) at any time in their life. Informed written consent was obtained from all subjects. While watching a video, subjects listened to three randomly presented

Mandarin stimuli resynthesized to differ only in f0: /mi1/ ‘to squint’, /mi2/ ‘bewilder’ and /mi3/ ‘rice’ (by convention, the number indicates tone or lexically meaningful pitch contour: Tone 1 = level tone, Tone 2 = rising tone and Tone 3 = dipping tone; see Supplementary Methods and Supplementary Fig. 1 online for details). Brainstem responses were collected using Scan 4.3 (Compumedics) with Ag–AgCl scalp electrodes. After f0 extraction (Supplementary Methods), we derived two primary measures of pitch tracking for each subject for each tone. First is the stimulus-to-response correlation (Pearson’s r between the f0 contour of the stimulus and the subject’s response contour), which indicates faithfulness of pitch tracking. Second is peak autocorrelation averaged over the entire response, which indicates robustness of neural phase-locking without making reference to the stimulus. In addition to these two primary pitch-tracking measures, we also considered the f0 amplitude of the FFR (which represents the average amount of spectral energy devoted to encoding the changing f0), the root-mean-square (RMS) amplitude of the FFR waveform, correlations between musical experience and pitch tracking, and subjects’ tone perception (behavioral) performances (see Supplementary Methods).

Each of the primary measures was entered into a 3 (tone)_2 (group) repeated measures ANOVA (for stimulus-to-response correlation, there was a main effect of group, $P < 0.015$, and tone, $P < 0.001$, but no significant interaction; for autocorrelation, there was a significant effect of tone, < o 0.001, but not of group, and a marginally significant interaction, Po0.08) followed by independent samples t-tests comparing group differences for each tone. The significance level was corrected for multiple comparisons following Bonferroni procedures. Overall, musicians showed more faithful representation of the stimulus f0 contours (Fig. 1, middle panels; Fig. 2a)

and more robust neural phase-locking (Fig. 1, bottom panels; Fig. 2b; see Supplementary Results online for details), particularly for the most complex contour (Tone 3). Musicians also showed stronger overall f0 amplitude and FFR RMS amplitude than nonmusicians (Fig. 1, top panels). Moreover, there was a significant positive correlation between the pitch tracking of the most complex contour and music experience (Fig. 3). Subjects also participated in tone identification and discrimination tasks, in which musicians showed significantly better identification ($t(18) = 3.664$, $P < 0.005$) and discrimination ($t(18) = 3.224$, $P < 0.005$). Subjects' performance on the discrimination task was significantly correlated with Tone 3 tracking (Pearson's $r = 0.434$, $P < 0.028$).

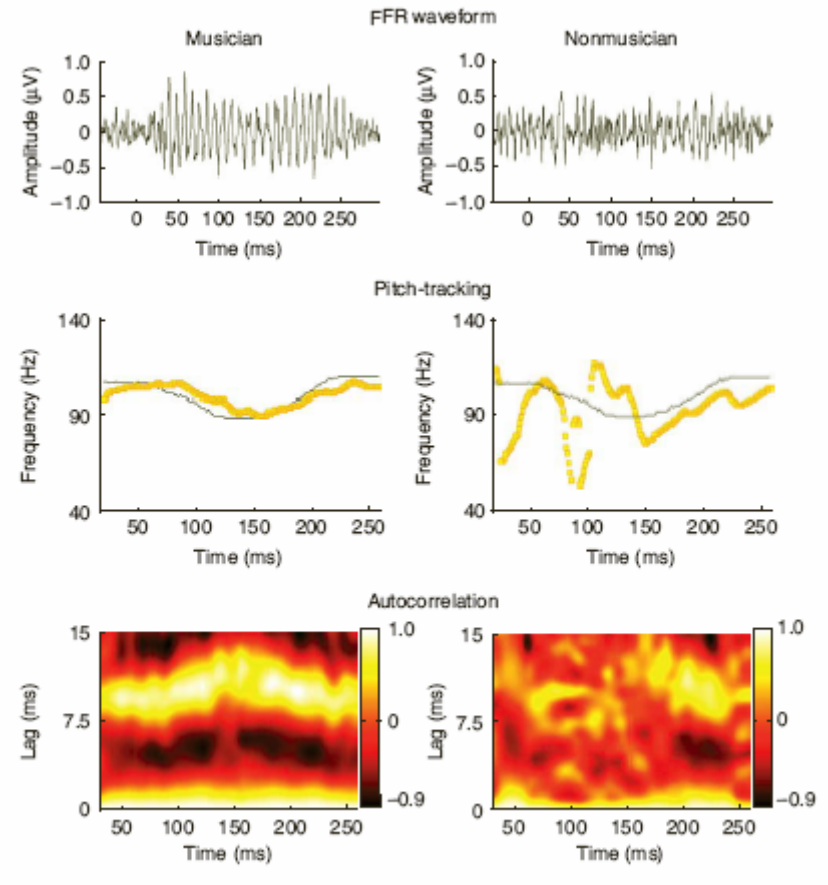

Fig. 1. Frequency following responses from selected subjects. Top, FFR waveforms from a musician (left) and nonmusician (right) elicited by a dipping pitch contour (Tone 3). Middle, trajectories (yellow line) of brainstem pitch tracking elicited by the same tone from the same subjects. The black line indicates the stimulus (expected) f0 contour. Bottom, autocorrelograms of the FFR waveforms. Color indicates the degree of correlation, with lighter colors indicating higher correlations. For the musician (left panel), the light band of color closely follows the inverse of the pitch contour of Tone 3 (frequency = 1/lag). In contrast, the nonmusician's autocorrelogram (right panel) is more diffuse and the highly correlated regions are not localized to the period of the f0 of the stimulus.

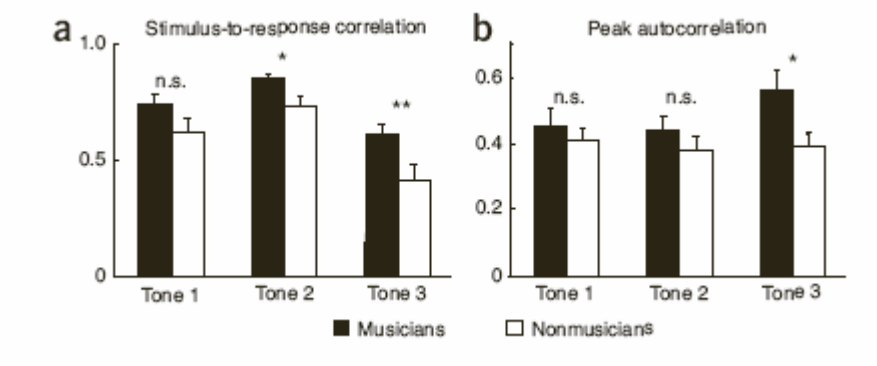

Fig. 2. Pitch tracking group results. (a,b) Mean stimulus-to-response correlation (a) and autocorrelation (b); black and white bars show averaged results from musicians and nonmusicians, respectively (error bars indicate one standard error; *P<0.03 and **P<0.016 based on independent samples t-tests).

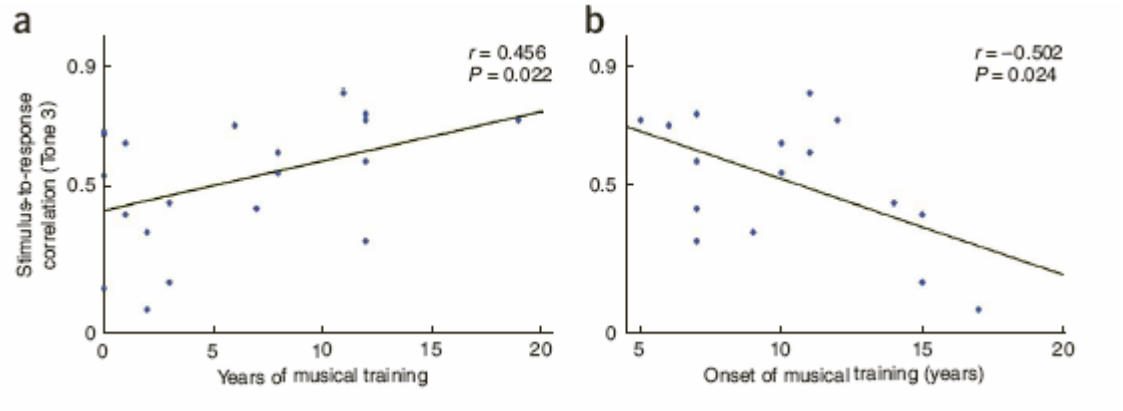

Fig. 3. Association between musical training and pitch tracking. (a,b) Correlations between Tone 3 tracking (stimulus-to-response correlation) and years of musicial training (a) and age at which musical training began (age onset; b).

Musical ability predicts the ability to produce and perceive the sound structures, but not grammatical or semantic structures, of a second language (Slevc, Miyake 2006). More specifically, musicians have an enhanced ability to learn lexical tones (Wong et al in press). Here, we found a plausible neurophysiological (subcortical) correlate of the effect of long-term musical training on speech (prosodic) encoding. Musicians have extensive experience using pitch information in the context of music, which requires both high cognitive demands and auditory acuity. This functional interplay is possibly mediated via feedback from the higher-level cortex to the inferior colliculus (made possible anatomically by the corticofugal pathway (Suga et al 2000)), such that accurate pitch information is relayed from subcortical structures to the neocortex to facilitate successful performance of cognitively demanding tasks. Cortical electrophysiology shows musical training to facilitate language processing in adults (Schon,

Magne, Besson 2004), and we are the first to show this effect in brainstem responses. Our line of reasoning is consistent with models of supervised perceptual learning involving changes in the weighting of perceptual dimensions as a result of feedback (Nosofsky 1986) and is also consistent with the reverse hierarchy theory of visual learning, which suggests that learning consists of an attention-driven, task-dependent 'backward' search for increased signal- to-noise ratio, especially for perceptual experts (Ahissar, Hochstein 2004). An important aspect of our results is that the musicians showed more robust and faithful neural encoding elicited by nonmusic stimuli, suggesting that corticofugal modulation is not entirely context specific. However, whether context-specific exposure still shapes the best response (for example, speech exposure effects on speech performance) requires further experimentation.

Although the current study provides evidence for the positive effect of long-term music exposure on speech (linguistic pitch) encoding at the brainstem, especially given the significant correlation between brainstem pitch tracking and music experience (in terms of both age of onset and years of musical training), we acknowledge that genetic differences between our musician and nonmusician groups could potentially account for the results. Moreover, our conclusion is limited by the small set of stimuli (Mandarin tones) used. However, because we have now established a robust effect and observed the pervasive impact of musical training on our nervous system, we believe a new line of research has been opened up, which would naturally involve more comprehensive and systematic investigations of musicians' and nonmusicians' responses to different simple and complex sounds.

In sum, we found more robust and faithful encoding of linguistic pitch information by musicians. Such encoding, arguably associated with increased musical pitch usage, may reflect a positive side effect of context-general corticofugal tuning of the afferent system, implying that long-term music-making may shape basic sensory circuitry. These results complement our existing knowledge of the brainstem's role in encoding speech (Johnson, Nicol, Kraus 2005) and frequency modulation (Gordon, O'Neill 2000) by demonstrating the interplay between music and speech, subcortical and cortical structures, and the impact of long term auditory experiences. Our findings have implications not only for biomedical sciences, but also for pedagogical principles and general social and educational policies.

ACKNOWLEDGMENTS

The authors thank J. Song, B. Williams, J. Alexander, A. Bradlow, T. Nicol, and T. Perrachione for their assistance in this research.

This work is supported by Northwestern University and the US National Institutes of Health grant R03HD051827 and R21DC007468 to P.W. and R01DC001510 to N.K, and National Science Foundation grant BCS-0544846 to N.K.

COMPETING INTERESTS STATEMENT The authors declare no competing financial interests.

1. Rauschecker, J.P. Curr. Opin. Neurobiol. 8, 516–521 (1998).

2. Patel, A.D. Nat. Neurosci. 6, 674–681 (2003).

3. Wong, P.C., Parsons, L.M., Martinez, M. & Diehl, R.L. J. Neurosci. 24, 9153–9160 (2004).

4. Wong, P.C.M., Perrachione, T.K. & Parrish, T.B. Hum. Brain Mapp. (in the press).

5. Russo, N.M., Nicol, T.G., Zecker, S.G., Hayes, E.A. & Kraus, N. Behav. Brain Res. 156, 95–103 (2005).

6. Greenberg, S., Marsh, J.T., Brown, W.S. & Smith, J.C. Hear. Res. 25, 91–114 (1987).

7. Krishnan, A., Xu, Y., Gandour, J. & Cariani, P. Cogn. Brain Res. 25, 161–168 (2005).

8. Li, W., Piech, V. & Gilbert, C.D. Nat. Neurosci. 7, 651–657 (2004).

9. Slevc, L.R. & Miyake, A. Psychol. Sci. 17, 675–681 (2006).

10. Suga, N., Gao, E., Zhang, Y., Ma, X. & Olsen, J.F. Proc. Natl. Acad. Sci. USA 97, 11807–11814 (2000).

11. Schon, D., Magne, C. & Besson, M. Psychophysiology 41, 341–349 (2004).

12. Nosofsky, R.M. J. Exp. Psychol. Gen. 115, 39–61 (1986).

13. Ahissar, M. & Hochstein, S. Trends Cogn. Sci. 8, 457–464 (2004).

14. Johnson, K.L., Nicol, T.G. & Kraus, N. Ear Hear. 26, 424–434 (2005).

15. Gordon, M. & O'Neill, W.E. J. Comp. Neurol. 426, 165–181 (2000).

Supplementary Methods

Subjects

Twenty subjects (11 females) participated in this study. None of the subjects had previous exposure to a tone language. Subjects were divided into two groups based on musical training. Amateur musicians were defined as instrumentalists having at least six years of continuous musical training (mean = 10.7 years) starting at or before the age of 12, in addition to currently playing their instrument. Nonmusicians were defined as having no more than three years of musical training (mean = 1.2 years) at any time in their life. Subjects' musical history information is summarized in Table S1. All subjects were right handed and reported no audiologic or neurologic deficits. All subjects had normal click-evoked auditory brainstem response latencies and normal hearing thresholds at or below 20 dB HL for octaves from 125 to 4000 Hz. The two subject groups did not differ in age or handedness scores.

Musician	Years of Training	Age Onset (Years)	Instrument (Lesson Type)
#1	6	6	Clarinet (Group)
#2	11	11	Piano (Private)
#3	12	12	Piano (Private)
#4	7	7	Piano (Private)
#5	8	11	Trumpet (Private)
#6	12	7	Piano (Private)
#7	12	7	Piano (Private)
#8	8	10	Piano (Private)
#9	12	7	Piano (Private)
#10	19	5	Piano (Private)
Mean	**10.7**	**8.3**	
Nonmusician			
#11	2	7	Keyboard (Private)
#12	2	10	Clarinet (Group)
#13	0	N/A	N/A
#14	1	15	Trombone (Private)
#15	3	15	Piano (Private)
#16	0	N/A	N/A
#17	1	10	Violin (Group)
#18	0	N/A	N/A
#19	0	N/A	N/A
#20	3	14	Piano (Private)
Mean	**1.2**	**13.33**	

Table S1. Subjects' musical history. Second and third columns indicate years of musical training and age at which musical training began (age onset), respectively. Mean age onset for nonmusicians was based on six subjects only.

Stimuli

A native speaker of Mandarin Chinese was asked to produce /mi/ with three Mandarin tones: /mi1/ 'to squint,' /mi2/ 'bewilder,' and /mi3/ 'rice' (by convention, the number indicates tone or lexically meaningful pitch contour: Tone 1 = level tone, Tone 2 = rising tone, and Tone 3 = dipping tone). Recording took place in a sound attenuated chamber using a SHURE SM58

microphone recorded at 44.1 kHz onto a Pentium IV PC. These original productions were then duration-normalized to 278.5 milliseconds (ms) using Praat (Boersma, Weenink 2004). Using Praat, the pitch (*f*0) contours of each of the original production were extracted and then superimposed onto the original Tone 1 (/mi1/) production using the Pitch-Synchronous Overlap and Add (PSOLA) method, which resulted in perceptually natural stimuli as judged by four native speakers of Mandarin. The stimuli, therefore, consisted of three instances of /mi/ (in three Mandarin tones) differing only in *f*0. These stimuli were RMS amplitude normalized using the software Level 16 (Tice, Carrell 1998). To accommodate the capabilities of our stimulus presentation software, the stimuli were resampled to 22.05 kHz. Fig. S1 shows the *f*0 contours of the three stimuli (*f*0 ranges: 140-172 Hz, 110-163 Hz, and 89-110 Hz, respectively). It is worth pointing out that we use the term "linguistic pitch" to describe these *f*0 contours because they were embedded in speech, not music. We realize that none of our subjects spoke a tone language and thus these *f*0 contours were not lexicalized. It is, therefore, likely that these *f*0 contours were interpreted as intonational tones, which also carry linguistic functions (Pierrehumbert 1980).

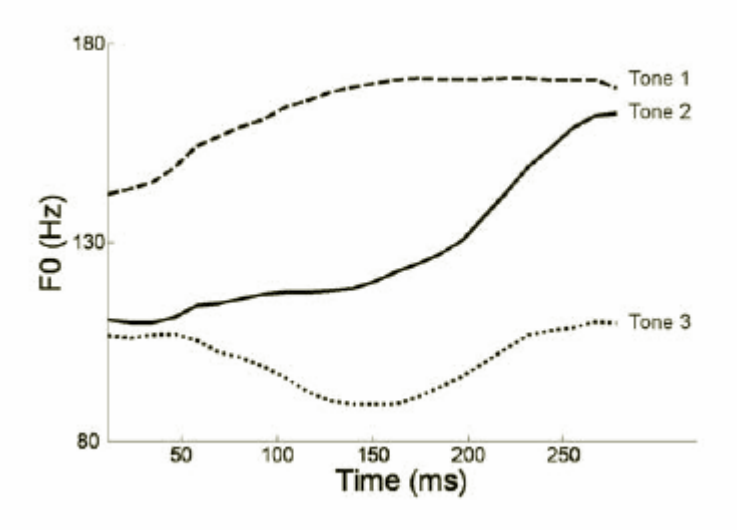

Fig. S1. *f*0 contours of the three stimuli (*f*0 ranges: 140-172 Hz, 110-163 Hz, and 89-110 Hz, respectively).

Physiologic (ERP) Recording Procedures

Physiologic recording procedures were similar to our published studies (e.g., Russo et al. 2004). During testing, subjects watched a videotape with the sound level set at < 40 dB SPL to facilitate a quiet yet wakeful state. Subjects listened to the video soundtrack (presented in free field) with the left ear unoccluded, while the stimuli were presented to the right ear through ER-3 ear inserts (Etymotic Research, Elk Grove Village, IL) at about 70 dB SPL (Stim, AUDCPT, Compumedics, El Paso, TX). The order of the three stimuli was randomized across subjects with a variable inter-stimulus interval between 71.50 and 104.84 ms. Responses were collected using Scan 4.3 (Compumedics, El Paso, TX) with Ag–AgCl scalp electrodes, differentially recorded from Cz (active) to ipsilateral earlobe (reference), with the forehead as ground. Two blocks of 1200 sweeps per block were collected at each polarity with a sampling rate of 20 kHz. Filtering, artifact rejection and averaging were performed offline using Scan 4.3. Responses were bandpass

filtered from 80-1000 Hz, 12 dB/octave, and trials with artifacts greater than 35 μV were rejected. Waveforms were averaged with a time window spanning 45 ms prior to the onset and 16.5 ms after the offset of the stimulus. Responses of alternating polarity were then added together to isolate the neural response by minimizing stimulus artifact and cochlear microphonic (Gorga, Abbas, Worthington 1985). For the purpose of calculating signal-to-noise ratios, a single waveform representing non-stimulus-evoked neural activity was created by averaging the neural activity 45 ms prior to stimulus onset.

Analysis Procedures

For each subject, we calculated two primary measures of FFR pitch-tracking: *stimulus-to-response correlation* and *autocorrelation*. These measures were derived using a sliding window analysis procedure in which 40-ms bins of the FFR were analyzed in the frequency and lag (autocorrelation) domains. The FFR was assumed to encompass the entire response beginning at time 1.1 ms, the transmission delay between the ER-3 transducer and ear insert. The 40-ms sliding window was shifted in 1 ms steps, to produce a total of 238 overlapping bins. A narrow-band spectrogram was calculated for each FFR bin by applying the Fast Fourier Transform (FFT) to windowed bins (Hanning window) of the signal. To increase spectral resolution, each time bin was zero-padded to 1 second before performing the FFT. The spectrogram gave an estimate of spectral energy over time and the *f*0 (pitch) contour was extracted from the spectrogram by finding the frequency with the largest spectral magnitude for each time bin. Spectral peaks that did not fall above the noise-floor were excluded as possible *f*0 candidates. Both *f*0 frequency and magnitude were recorded for each time bin, and the *f*0 *amplitude* measure was calculated as the

average magnitude across bins. The same short-term spectral analysis procedure was applied to the stimulus waveforms to calculate the degree of similarity (*stimulus-to-response correlation*) between the stimulus and response *f*0 contours, defined as the Pearson's correlation coefficient (*r*) between the stimulus and response *f*0 contours. This measure represents both the strength and direction of the linear relationship between to two signals. The second measure of pitch-tracking, *autocorrelation*, was derived using a pitch detection short-term autocorrelation method (Boersma, 1993). Each of the 238 time bins was cross-correlated with itself to determine how well the bin matched a time-shifted version of itself. The maximum (peak) autocorrelation value (expressed as a value between 0 and 1) was recorded for each bin, with higher values indicating more periodic time frames. The *autocorrelation* pitch tracking measure was calculated by averaging the autocorrelation peaks (*r*-values) from the 238 bins for each tone for each subject. Running-autocorrelograms (lag versus time) (see Krishnan et al.2005) were calculated as a means of visualizing and quantifying periodicity and pitch strength variation over the course of the response. In the pitch-tracking and autocorrelation plots (Fig. 1, middle and bottom panels), the time indicated on the *x*-axis refers to the midpoint of each 40-ms time bin analyzed. For example, the *f*0 extracted from the first FFR time bin (1.1 ms - 40.1 ms) is plotted at time 21.1 ms. We also measured the RMS (Root-Mean-Square) amplitude of the FFR waveform, which is the magnitude of neural activation over the entire FFR period (1.1 –295 ms). This measure takes both negative and positive peaks into consideration. This *FFR RMS amplitude* is driven largely by the amplitude of the *f*0 (a description of the *f*0 *amplitude* calculation is provided above*).* If a subject has robust pitch-tracking, the largest peaks in the response waveform will fall at the period of the *f*0. In addition, to quantitatively consider the proportion of the *f*0 amplitude relative

to the overall *FFR RMS amplitude*, we calculated *f*0-*FFR proportion*, which is the average *f*0 amplitude divided by the total RMS amplitude. The use of multiple pitch-tracking measures allows us to more comprehensively observe and quantify pitch encoding differences between the two groups. All pitch-tracking analyses were performed using routines coded in Matlab 7.4.1 (Mathworks, Natick, MA , 2005).

Behavioral Testing (Tone Identification and Discrimination)

Subjects also participated in two behavioral experiments designed to test their ability to identify and discriminate Mandarin tones. The stimuli and procedures, summarized briefly here, were essentially identical to Alexander, Wong, and Bradlow 2005). Stimuli consisted of twenty monosyllabic Mandarin Chinese words. The five syllables /bu/, /di/, /lu/, /ma/, /mi/ were each produced in citation form with the four tones (level, rising, dipping, and falling) of Mandarin. Talkers consisted of two male and two female native speakers of Mandarin Chinese. Subjects participated in these two experiments after task familiarization. In tone identification, subjects matched the auditory stimulus with visually presented arrows depicting the pitch trajectory. In tone discrimination, subjects made a same-different judgment on the pitch patterns of stimulus pairs.

REFERENCES

1. Boersma, P. & Weenink, D. (2004)

2. Tice, R. & Carrell, T. D. (1998).

3. Pierrehumbert, J. B. (MIT, Cambridge, 1980)

4. Russo, N., Musacchia, G., Nicol, T., Zecker, S. & Kraus, N. Brainstem responses to speech syllables. *Clinical Neurophysiology* **115**, 2021-2030 (2004).

5. Gorga, M. P., Abbas, P. J. & Worthington, D. W. in *The Auditory Brainstem Response* (ed. Jacobson, J. T.) 49-62 (College-Hill Press, San Diego, 1985).

6. Boersma, P. in *Proceedings of Institute of Phonetic Sciences* 97-110 (Amsterdam,1993).

7. Krishnan, A., Xu, Y., Gandour, J. & Cariani, P. Encoding of pitch in the human brainstem is sensitive to language experience. *Cogn Brain Res* **25**, 161-8 (2005).

8. Alexander, J., Wong, P. C. M. & Bradlow, A. in *Proceedings of Interspeech 2005 - Eurospeech - 9th European Conference on Speech Communication and Technology* (2005).

ON THE RELATIONSHIP BETWEEN SPEECH AND NONSPEECH EVOKED AUDITORY BRAINSTEM RESPONSES

Published in *Journal of Audiology and Neurootology*

Song JH[a,b], Banai K[a,b,e], Russo NM[a,b,e], Kraus N[a-e]

[a]Auditory Neuroscience Laboratory, Departments of [b]Communication Sciences; [c]Neurobiology and Physiology; [d]Otolaryngology and [e]Northwestern Institute of Neuroscience, Northwestern University, Evanston, IL 60208.

Abstract

Auditory brainstem response (ABR) reflects activation of the neural generators along the ascending auditory pathway when a sound is heard. In this study, we explored the relationship between brainstem encoding of click and speech signals in normal learning children and in those with language-based learning problems. To that end, ABR was recorded from both types of stimuli. We found that the normal pattern of correlation between click and speech-evoked ABRs was disrupted when speech-evoked ABRs were delayed. Thus, delayed responses to speech were not indicative of clinically abnormal responses to clicks. We conclude that these two responses reflect largely separate neural processes and that only processes involved in encoding complex signals such as speech are impaired in children with learning problems.

Key words: auditory processing, ABR, speech, learning disability

Introduction

Despite decades of intensive research, the biological underpinnings of language-based learning disabilities, affecting approximately ten percent of school-aged children [Torgesen, 1991], are not well understood. As a consequence, objective and early diagnosis of learning disabilities, which is desirable from a therapeutic and educational standpoint, remains a complicated matter.

The auditory brainstem response (ABR) is generated by synchronous firing of structures along the ascending auditory pathway, which include the auditory nerve, cochlear nuclei, superior olivary nuclei, lateral lemnisci, and inferior colliculi [Møller and Jannetta, 1985]. The ABR is ideally suited for evaluating difficult-to-test patients because it is a passively elicited neurophysiological response to auditory stimuli and does not require the patient to actively attend or respond to the stimulus. The click-evoked ABR is used widely by clinicians when evaluating hearing and the integrity of the auditory brainstem in certain populations, such as infants or neurologically impaired patients [Starr and Don, 1988]. A normal click-evoked response latency is defined as occurring within two standard deviations of the normal population [Hall, 1992]. Specifically, wave V latency to a 80 dB nHL click, typically occurring 6.25 ms from stimulus onset for infants [Gorga et al., 1989] and 5.47 ms for adults [Hood, 1998], is extensively used in clinical settings. Thus, the click-evoked auditory brainstem response (ABR) has proven to be a valuable measure in evaluating auditory function, even helping to distinguish between sensorineural and conductive hearing loss [Hall, 1992; Hood, 1998; Jacobson, 1985].

In addition to clicks, ABRs can be evoked using a wide array of stimuli, including pure tones,

masked tones [Marler and Champlin, 2005], and speech sounds [Krishnan, 2002; Russo et al., 2004]. The speech-evoked ABR can be divided into transient and sustained portions, specifically the onset response and the frequency-following response (FFR) [Johnson et al., 2005; Kraus and Nicol, 2005]. Onset responses are transient, similar to click-evoked ABR, with peak durations lasting tenths of milliseconds. Although the FFR is an important feature of speech-evoked ABR, it is not further explored here; rather, the relationship between the onset responses to click and speech stimuli is the primary focus of this study.

The relationship between the click-evoked and speech-evoked ABRs is not clear. Previous studies have typically documented normal click-evoked ABR responses in children diagnosed with learning disability [Grøntved et al., 1988a; Grøntved et al., 1988b; Jerger et al., 1987; Jirsa, 2001; Lauter and Wood, 1993; Mason and Mellor, 1984; McAnally and Stein, 1997; Purdy et al., 2002; Tait et al., 1983]. These findings have been taken to indicate that the structural integrity of the ascending auditory pathway in children with a learning disability is intact. However, when measured by psychophysical tasks, approximately thirty percent of all individuals with a learning disability suffer from poor auditory processing [Ahissar et al., 2000; Amitay et al., 2002; Banai and Ahissar, 2004; Menell et al., 1999; Ramus et al., 2003; Tallal, 1980]. These studies, combined with studies of cortical evoked responses [Baldeweg et al., 1999; Kraus et al., 1996; Lachmann et al., 2005; Nagarajan et al., 1999; Paul et al., 2005], contributed to the view that inasmuch as auditory processing deficits are relevant to the etiology and diagnosis of learning disabilities, the physiological deficit has cortical origins [Heim and Keil, 2004].

More recent studies, however, suggest a sub-cortical origin for learning disabilities. In these studies, a subset of children with learning disabilities show abnormal neural encoding of a speech syllable at the level of the brainstem (speech-evoked ABRs) [Cunningham et al., 2001; King et al., 2002; Wible et al., 2004; Wible et al., 2005; Johnson et al., 2005]. In particular, abnormal onset responses of the speech-evoked ABRs characterize approximately thirty percent of the learning impaired children [Banai et al., 2005]. Although the timing of the click-evoked response is within normal limits, the onset of the speech-evoked ABR appears to be delayed and less robustly synchronized in these children, leaving the nature of the relationships between these two measures unclear.

The current study investigates the relationship between click and speech-evoked ABR as recorded in children both regarded as typically developing and those clinically diagnosed with a learning problem. Since both the response to a click stimulus and the onset ABR to speech occur in a similar time frame, and are thought to originate from similar locations, a relationship between the two may reflect a similar type of neural processing. In other words, the finding that both nonspeech and speech auditory stimuli (i.e. click and /da/) elicit analogous brainstem responses within the first 10 ms from the onset of the stimulus would suggest that these sounds activate a similar set of neural operations as it ascends along the auditory brainstem pathway. Thus, we asked, are these two measures related to each other, and if so, is an abnormal speech-evoked ABR indicative of a clinically abnormal click-evoked ABR? Based on previous studies, it was expected that the click-evoked ABRs of the majority of children with learning problems would be clinically normal, irrespective of their speech-evoked ABR, although it was possible

that their responses would be delayed but still within clinical norms. Furthermore, it was expected that among children with delayed speech-evoked ABRs, the normal pattern of correlation between the speech and click evoked measures would be altered, reflecting the different processing (normal vs. impaired) of the two types of stimuli. To this end, the relationship between click and speech-evoked ABR in normally developing children and those with learning disabilities was systematically examined to determine if having a delayed ABR to speech was predictive of neurophysiologic timing differences to click stimuli.

Methods

Participants

Two hundred and thirty four native English-speaking children (8-12 years old) participated in this study. All participants had normal hearing thresholds at or below 20 dB HL for octaves from 500 to 4000 Hz and I.Q. scores ≥ 85 as measured with the Brief Cognitive Scale [Woodcock and Johnson, 1989] or the Test of Nonverbal Intelligence (TONI-3) [Brown et al., 1997]. Consent and assent were obtained from the parents (or legal guardians) and the children involved in the study. The Institutional Review Board of Northwestern University approved all research.

Two groups participated in this study. One group comprised 119 children diagnosed with a learning problem (LP) by outside professionals (clinical psychologists, school psychologists, neurologists, etc.) and verified by their performance on study-internal standardized measures of

learning and academic achievement described below. The second group comprised 115 normal learning (NL) children who were never diagnosed with a learning problem.

Study-internal measures

A psycho-educational test battery given to all participants included subtests taken from Woodcock-Johnson Revised [Woodcock and Johnson, 1989]. These subtests were Auditory Processing (Incomplete words and Sound Blending), Listening Comprehension, Memory for Words, Cross-out, and Word Attack. Additionally, reading and spelling skills were assessed by using subtests from Wide Range Achievement Test-3 [Wilkinson, 1993] and phonological skills were assessed by using subtests taken from the Comprehensive Test of Phonological Processing (CTOPP) [Wagner et al., 1999]. These subtests were Elision, Phoneme Reversal, and Segmenting Nonwords.

Stimulus and recording parameters

Auditory brainstem responses were elicited by an acoustic click and a speech syllable, /da/, and both brainstem responses were collected in the same manner and during the same recording session. Responses were recorded from Ag-AgCl electrodes, with contact impedance of less than 5 kΩ, positioned centrally on the scalp, at Cz, behind the right ear lobe (reference) and on the forehead (ground). Stimuli were presented into the right ear at 80.3 dB SPL through insert earphones (ER-3, Etymotic Research, Elk Grove Village, IL). The sampling rate was 20,000 Hz and responses were online bandpassed filtered from 100 to 2000 Hz, 6 dB/octave. Trials with eye-blinks or other motion artifacts greater than 35 μV were rejected. During testing, the

children watched a videotape with the sound level set at less than 40 dB SPL in free field so they could hear it in the non-test ear.

For the click-evoked response, the stimuli were 100 μs clicks presented at a rate of 31.1 Hz. A click is a brief square wave with broad spectral content (see Figure 1). Three blocks of 1000 sweeps each were collected both in quiet and ipsilateral white Gaussian noise (+5 dB SNR) conditions. Waveforms were averaged online in Neuroscan (Compumedics, El Paso, TX). The recording window was 20 ms starting 5 ms prior to stimulus onset.

For the speech-evoked (da) response, stimuli were presented at a rate of 11.1 Hz. The 40 ms /da/ stimulus was a five-formant synthesized stimulus [Klatt, 1980] and contained an initial 10 ms burst with frequencies of which centered around the beginning frequencies of formants 3 to 5 in the range of 2580 to 4500 Hz (see Figure 1). Three blocks of 1000 sweeps each were collected in quiet. Responses of alternating polarity were added together to isolate the neural response by minimizing stimulus artifact and cochlear microphonic [Gorga et al., 1985]. Waveforms were averaged online in Neuroscan. The recording window was 70 ms starting 10 ms prior to stimulus onset.

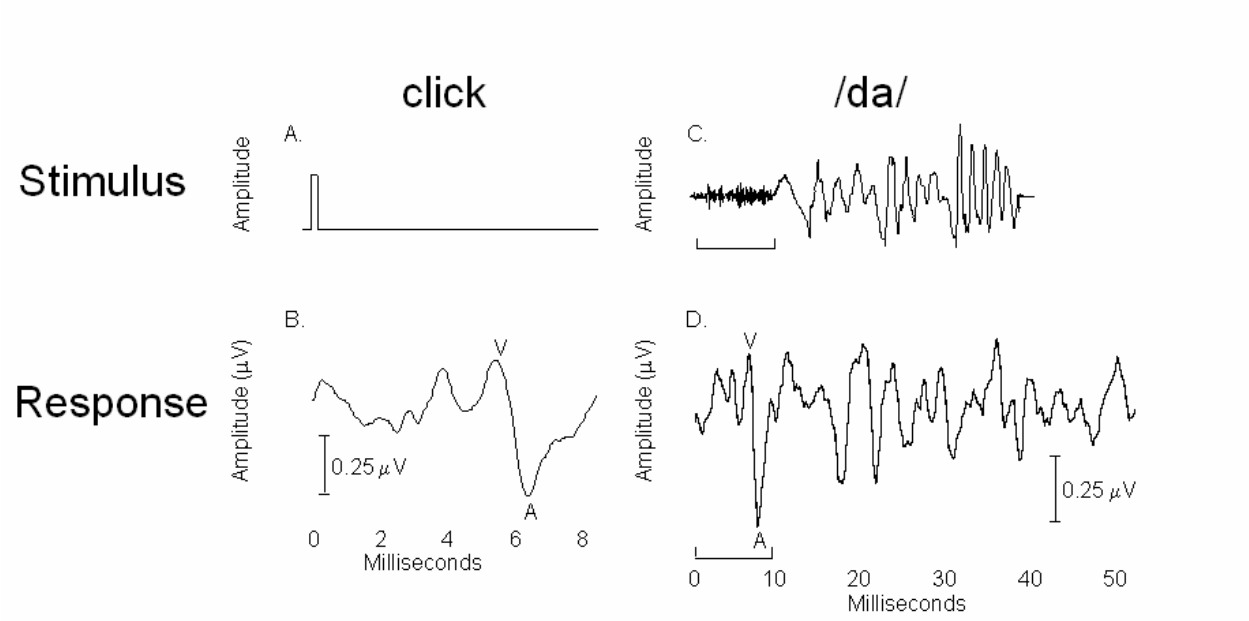

Fig. 1. Stimulus waveforms for the click and /da/ stimuli (top) and their corresponding grand average response waveforms in quiet (bottom). A. The click is a brief sound with a rapid onset and duration and broad range of frequencies. B. The click-evoked ABR normally consists of characteristic peaks (i.e. I, III and V) in the waveform at predictable latencies; the most robust positive peak being wave V which is followed immediately by its negative trough (wave A). C. The /da/ stimulus is a synthesized speech-like sound and contains the onset burst frication of the third, fourth and fifth formant frequencies during the first 10 ms, followed by 30 ms of the first and second formant transitions which stops promptly before the sustained vowel portion (see Johnson et al. (2005) for further detail of stimulus). D. The onset of the speech-evoked ABR includes a positive peak (wave V) followed immediately by its negative trough (wave A). The onset portion of the /da/ stimulus and response is bracketed from the entire waveform and reflects its transient quality. The sustained activity beginning at approximately 18 ms is the frequency-following response to the vowel portion of the /da/.

Data analysis

Experienced observers manually marked wave V and its negative trough (wave A) latencies of click-evoked responses, recorded both in quiet and noise, blind to participants' identity and diagnostic category. The voltage difference between these two peaks was used as response amplitude. The response measures of NL and LP groups were compared. Repeated measures analysis of variance (ANOVA) was used for statistical analysis of correlation and amplitude measurements.

For the speech-evoked ABR, response measures that were considered for evaluation were wave V latency, wave A latency, VA interpeak duration, and VA interpeak slope. Experienced observers manually marked wave III, V and A latencies from responses recorded in quiet.

Normal and delayed ABR defined

As previously mentioned in the introduction, for clinical application, values that exceed 2 standard deviations of the normal population are considered abnormal [Hall, 1992], thus this criterion was adopted for this study. Since the average click-evoked wave V latency of the normal group in our study was 5.87 ± 0.30 ms, abnormal values were defined as those exceeding 6.47 ms. The participants' onset ABR to /da/ was considered abnormal if at least two of the aforementioned measures were beyond 1.5 SD of the normative values or at least 1 measure was beyond 2.0 SD of the normative values [Banai et al., 2005].

Results

Click-evoked ABR

Latency and amplitude values of the click-evoked ABRs for NL and LPs are displayed in Table 1 and averaged click-evoked ABRs recorded in quiet for each participant group are shown in Figure 2**.** A 2-factor repeated measures ANOVA with group (NL vs. LD) as the between subjects factor and condition (quiet vs. noise) as the within subjects factor performed separately for latency and amplitude values revealed that these values did not significantly differ between NLs and LDs in either quite or noise (for latency values: F_{group}=0.002, p=0.964; $F_{condition}$ = 210, p=0.000; $F_{interaction}$ = 1.857, p=0.174 ;for amplitude values: F_{group}=2.44, p=0.12; $F_{condition}$ = 166, p = 0.000; $F_{interaction}$ =0.232, p=0.631) . Thus, background noise introduced a delay in latency and a reduction in amplitude; however, the influence of noise was similar in both groups.

	Quiet		**Noise**	
	V Lat (ms)	**VA Amp (μV)**	**V Lat (ms)**	**VA Amp (μV)**
NL (n=115)	5.87 (0.30)	0.37 (0.37)	6.14 (.0.29)	0.19 (0.28)
LP (n=119)	5.85 (0.26)	0.31(0.37)	6.13 (0.36)	0.20 (0.25)

Table 1. NL and LP subgroup means (± 1SD) of wave V latency and amplitude click-evoked measures in quiet and noise conditions.

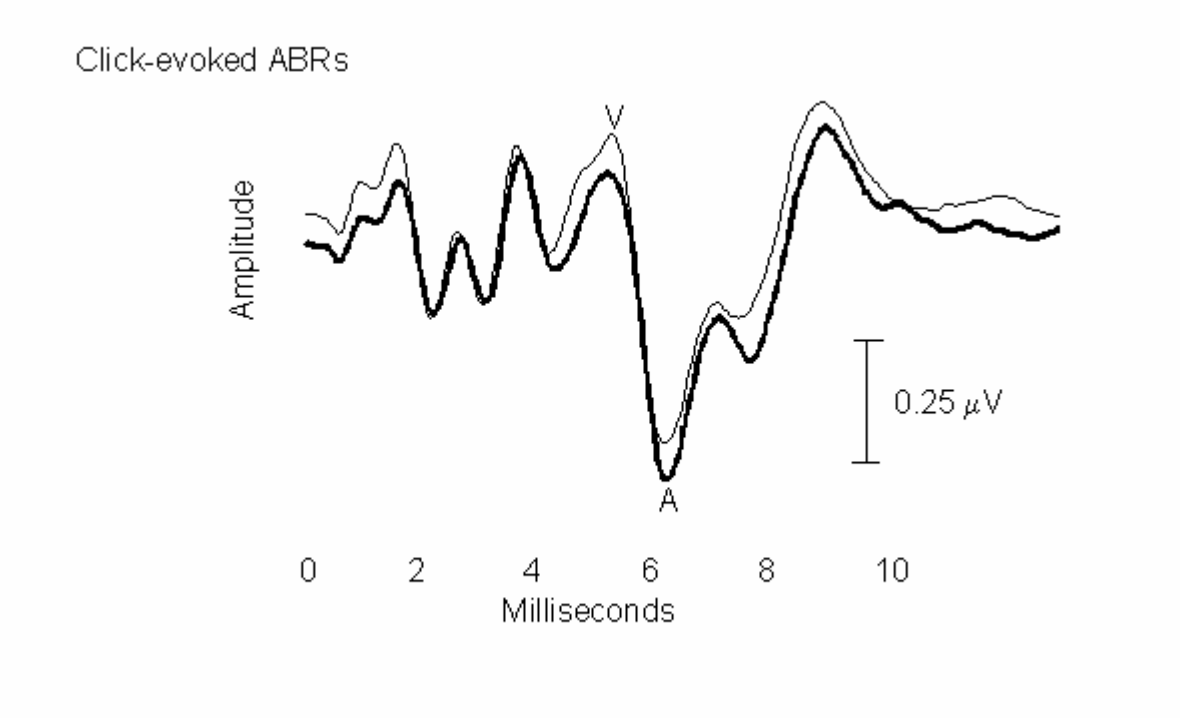

Fig. 2. Comparison of grand averaged click-evoked ABRs recorded in quiet between normal learning children (thin line) and those diagnosed with a learning problem (thick line). No significant latency or amplitude differences were found between NL and LP children in response to clicks.

When the correlation between click measures in quiet and noise in both groups were examined, the NLs and LPs showed comparable effects of background noise. Within both NL and LP groups, each showed similarly strong and moderate correlations between quiet and noise conditions in latency and amplitude respectively, as shown in Table 2.

	Wave V	R
NL (n=115)	Latency	0.611
	Amplitude	0.372
LP (n=119)	Latency	0.600
	Amplitude	0.353

Table 2. Correlations between quiet and noise conditions of click-evoked wave V latency and amplitude. All correlations are significant at $p<0.001$

Speech-evoked ABR

Further analysis of ABR data involved assessment of speech-evoked responses in our participants. Upon evaluation of the onset-ABR measures to /da/, 183 children (97 NL, 86 LP) exhibited a normal response to the /da/ stimulus and 51 children (18 NL, 33 LP) exhibited abnormal response (see Methods for definition) to the same stimulus (see Figure 3). Among the 18 NLs with abnormal speech-evoked ABR, nine children were suspected to have a possible learning problem based on parental report or study-internal measures; however, the lack of formal diagnoses during the time of our testing precluded their inclusion in the LP group.

The waveforms of the two resulting groups are shown in Figure 3. As expected, the waveform of the abnormal speech group shows a characteristic delay between waves V and A and a reduced transition slope. In addition to the VA complex measures, this grouping of speech-onset responses also showed a delayed wave III latency in response to /da/ ($t=121.36$, $p=0.000$). The

mean latencies and standard deviations of wave III of normal and abnormal speech-onset group were 4.84 ±0.25 ms and 5.07 ±0.35 ms respectively.

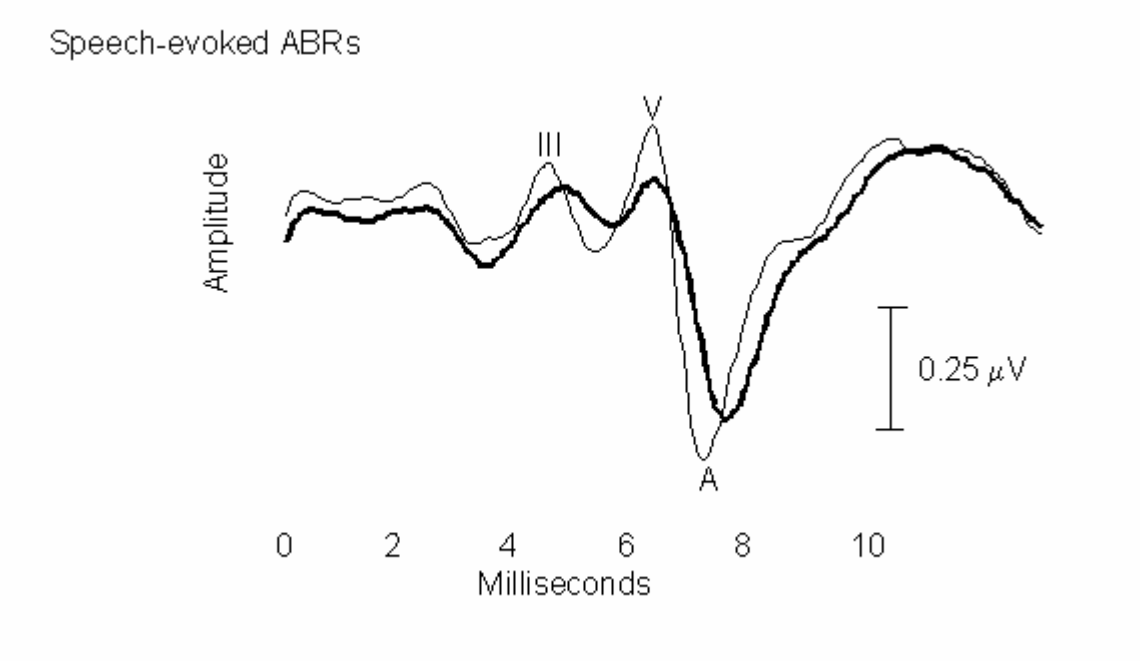

Fig. 3. Comparison of grand averaged speech-evoked onset response (V, A) between children with normal (thin) and abnormal (thick) speech-evoked onset response. In addition to the VA complex measures which are impaired by definition, this grouping also revealed delayed wave III latency in response to /da/.

Click vs. speech-evoked onset responses

The relationship between click and speech-evoked measures was also examined (see Figure 4). In our entire test population, latency of click wave V correlated moderately, but significantly, with the latencies of speech onset response V and A ($r=0.47$, $p<0.0001$; $r=0.44$, $p<0.0001$, respectively). There was a weak correlation between click wave V latency and VA slope measure for speech ($r=-0.21$, $p=0.001$). On the other hand, click wave V latency did not

correlate with VA duration measure for speech (r=0.07, p=0.28). These findings suggest that while there may be some shared processing reflected in the click and speech onset latency measures, there is also a separate component unique to the processing of more complex auditory signals, such as speech. This pattern of correlation was almost identical in the NL and LP groups, suggesting that as a rule, the normal pattern of relationship between the encoding of click and speech stimuli at the brainstem level is not disrupted in individuals with learning problems.

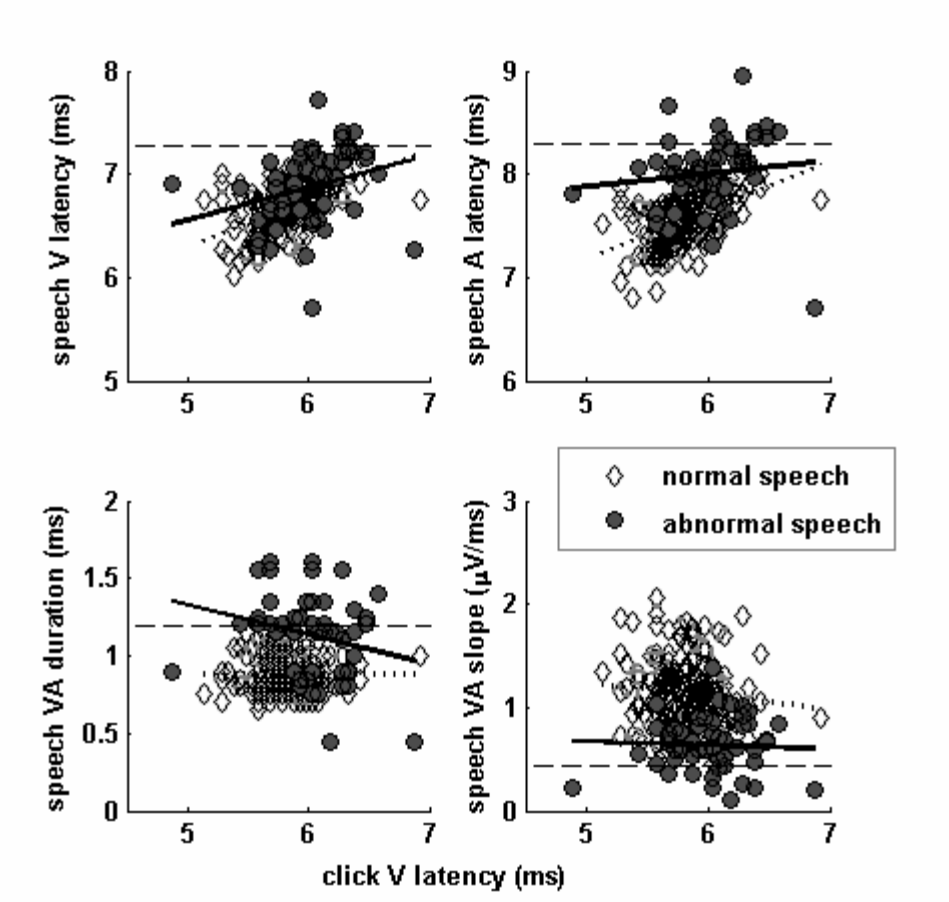

Fig. 4. Speech ABR measures (V latency, A latency, VA duration, and the absolute value of VA slope) as a function of click V latency. Diamonds represent participants who exhibit within normal speech ABR and filled circles represent those with speech-ABRs that are abnormal on

any one of the four measures. The dashed horizontal lines indicate the normal limit. Dotted lines depict the linear fit of the click and speech measures in the normal speech-ABR group, black lines depict the linear fit among delayed speech-ABR participants.

When comparing individuals with normal and abnormal speech-ABRs, a significant difference in click wave V latency was observed (5.82 +/- 0.25 ms vs. 6.01 +/- 0.34 ms, respectively, t = 3.847; p =0.000) raising the possibility that delays in both reflect a similar process. However, the variance in speech-ABR cannot be fully accounted for by variance in click wave V latencies. An ANCOVA with click-evoked wave V latency as the covariate and speech-evoked wave V latency as the dependent variable showed that the difference in speech-evoked wave V latency remained significant even when the click-evoked wave V latency was adjusted for variability (F=49.463, p=0.000). This finding suggests that the difference in click-evoked wave V latency cannot fully account for the difference in speech-evoked wave V latency.

Furthermore, the click-speech correlations were driven by those subjects with a normal /da/ response. When divided between participants with normal (n=183) and abnormal speech-evoked responses (n=51), the correlations between the speech and click measures were significantly reduced in individuals with delayed speech-evoked ABRs compared to individuals with normal speech-evoked ABRs (see Table 3).

	V latency	A latency	VA slope	VA duration
Normal (n=183)	0.51 (<0.001)	0.49 (<0.001)	-0.10 (ns)	0.00
Delayed (n=51)	0.27 (.051)	0.11 (ns)	0.04 (ns)	-0.24 (ns)
Group difference	z=1.93, p<0.05	z=2.62, p<0.01	Z=0.37 (ns)	z=1.51 (ns)

Table 3. Pearson correlation (p value) between click wave V latency and speech-evoked onset ABR measures in individuals with normal and abnormal speech-evoked ABRs. Group difference shows the comparison of correlation coefficients between normal and delayed speech ABR measures.

While by definition all subjects in the abnormal speech group had VA measures that would have been considered abnormal in clinical terms (i.e. delayed or reduced by 2 standard deviations), these same subjects typically had clinically normal click-evoked ABRs. Moreover, the proportion of children with normal click-evoked ABRs did not significantly differ as a function of speech-evoked ABR (normal vs. delayed z=-0.0025, ns) where 99.45% of the participants exhibited normal click and normal speech-evoked ABR and 96.08% exhibited normal click and abnormal speech-evoked ABR. Thus, the speech-evoked ABR provides additional, potentially clinically significant information about sound encoding at the individual level, not provided by the click-evoked measure. These findings indicate that even though some aspects of the click and speech-evoked ABRs are correlated, each provides a separate type of information and thus, a delayed speech-evoked measure does not necessarily predict a delayed click. Both click and

speech-evoked responses should be evaluated in order to ascertain a broader knowledge of auditory processing ability.

Discussion

Objectively identifying children at risk for learning problems at an early stage in development constitutes an important advance in their diagnosis and prospects. Children with learning problems demonstrated wave V latencies within normal limits in response to a click stimulus presented not only in quiet, but also in the presence of background noise. Reinforcing previous findings, this large-scale study demonstrated that normal hearing children with learning problems almost always have a normal auditory brainstem response to click stimuli in quiet and noisy environments. Furthermore, these findings demonstrated a comparable influence of noise in the encoding of click stimuli in both the normal learning and learning impaired children at the level of the brainstem. Background noise distorted wave V latency and amplitude similarly for normal learning and learning impaired children in that the timing of wave V latency was delayed and the amplitude of wave V was reduced. Thus, these findings suggest that abnormal processing of brief stimuli, such as a click, in either quiet or noise is unlikely to play a role in the diagnosis of learning disability.

On the other hand, the brainstem response to speech has proven to be a mechanism for understanding the neural bases of normal attention-independent auditory function [Johnson et al., 2005; Kraus and Nicol, 2005; Russo et al., 2004]. Because a speech signal provides different acoustic information than a click (i.e. speech syllables are longer and contain less high frequency

information compared to clicks), it provides additional information about neural encoding at the brainstem level and it uncovers abnormal encoding in approximately thirty percent of children with learning problems [Banai et al., 2005; Cunningham et al., 2001; Johnson et al., 2005; King et al., 2002; Wible et al., 2004]. Furthermore, the integrity of processing of speech at the level of the brainstem is highly related to the robustness of the cortical response in noise [Wible et al., 2005] suggesting that brainstem processing of these two types of signals may be related to cortical processing in different ways.

When the normal and abnormal speech-evoked ABR waveforms were compared, in addition to the expected differences in wave V and A, we have interestingly observed a difference in wave III. A difference in wave III latency to speech suggests a disruption of an earlier component in the response occurring lower in the brainstem. Wave III is less readily identified compared to waves V and A; therefore, it is probably not clinically useful. Further analysis to examine these differences in speech-onset responses will be explored in future studies. Since children with abnormal speech-evoked ABRs tend to show delayed wave Vs in response to clicks, it may be claimed that differences in the response to speech may be accounted for by the delayed response to clicks. Our statistical analysis has shown that this is not the case since differences in speech-evoked parameters remained significant even when click latencies were controlled. Furthermore, since click latencies were within clinical norms for the large majority (96%, see results) of children with abnormal speech responses, they cannot be used to group individual subjects as 'abnormal'. Thus, speech-evoked ABRs possibly provide additional diagnostic information at the individual level that is hard to obtain using only clicks.

The speech-evoked response in noise was not reported here as earlier work [Banai et al., 2005; Cunningham et al., 2001; Johnson et al., 2005 King et al., 2002; Russo et al. 2004; Wible et al., 2004] demonstrated that the /da/-evoked response in quiet sufficiently provides a means to objectively identify children at risk for learning problems; moreover, Russo et al. (2004) reported that the onset waves of the /da/-evoked response in noise are not reliably identifiable. The current finding of unimpaired click-evoked responses in quiet in the LP group gave a strong indication that other factors such as stimulus complexity were responsible for the abnormal LP speech-evoked ABR.

The auditory encoding differences between the click and speech stimuli may be derived from an examination of the differences in their acoustical structures. The click stimulus is a non-periodic, relatively simple sound that is short in duration, but whose bandwidth contains a broad range of frequencies. Conversely, consonant-vowel speech syllables, such as the /da/ used in this study, begin with rapid, relatively low amplitude transient onset features that may be especially vulnerable to disruption by background noise [Brandt and Rosen, 1980]. The vowel that follows the consonant is a sustained periodic signal that is much louder than the consonant. Thus, this higher amplitude, longer portion of the stimulus may actually mask the brief consonant onset critical for eliciting the onset portion of the speech-evoked ABR. This effect may be especially pronounced in the learning impaired population which is known to show larger perceptual effects of backwards masking compared to normal learning children [Johnson et al., 2004; Wright et al., 1997]. Moreover, children with abnormal speech-evoked ABR are more likely to have increased

backward masking compared to those with normal encoding of speech at the brainstem [Johnson et al., 2004]. Recent findings showing increased physiological effects of backward masking in children with specific language impairment [Marler and Champlin, 2005] further suggest that deficient neural mechanisms handling backward masking may partially explain the differences between the click (an unmasked stimulus) and speech (consonant onset masked by the steady state vowel) evoked responses observed here. Marler and Champlin [2005] measured ABRs in a group of children with language problems in two conditions. In the unmasked condition, in which ABRs were evoked using a tone, responses were normal; however, when the tone was immediately followed by a masker, wave V latency was significantly delayed. This interpretation should be further explored by a systematic manipulation of the temporal position of the signal and masker.

Another feasible explanation for the differences observed between the encoding of the click and speech signals involves possible differences in neural populations recruited during the encoding click and speech auditory stimuli. Our statistical analysis (both correlational findings and ANCOVA) indicated that while the latency of click and speech-evoked responses may share at least some common neural processing, variations in the latency of the speech-evoked waves cannot be accounted for entirely by the same physiological processes underlying the processing of the click stimulus. Thus, these findings suggest that the encoding of speech sounds may recruit processes that are not present in the encoding of the click stimuli. The processing of complex features present in speech, such as the onset and formant structures of speech sounds, may indeed require separate processes in order to encode the sound accurately; these processes

may be compromised in children with delayed speech-evoked ABRs. Our findings that the pattern of correlation between click and speech evoked responses differed as a function of the precision of speech encoding, together with the fact that the compromised processing of a speech signal is not an indication of abnormal processing of a click stimulus, reinforce this notion.

On the other hand, the differences between encoding of the click and speech stimuli also suggest that abnormal speech-evoked ABR may likely be based on differences in synchronization of response generators in the brainstem. Thus, if a neural system is more sensitive to effects of desynchronization, this increased susceptibility will become apparent in response to the speech stimulus which is longer in duration and has a more gradual onset compared to the click. The abnormal latency in response to the /da/ stimuli in these children may, in fact, reflect this diminished synchronization of ABR wave V-A generators through greater dispersion of latencies of neural activity that is contributing to the response [Wible et al., 2004]. Further studies, probably in an animal model, is needed to determine if the differences in encoding arise from distinct neural populations that are recruited to encode the complex features of the speech stimuli or from the differential influence of different stimuli on the same neural population.

In summary, in this study we have demonstrated the relative independence of brainstem encoding of a brief, broad spectrum click and a longer duration, harmonically and temporally complex speech syllable. At an early level of the brainstem, processing of acoustic input is differentiated based on the acoustic properties of the stimulus. This distinction suggests that evaluating each of these responses may have a unique clinical role. While normal click-evoked

ABRs are an indication of the integrity of the cochlea and the ascending auditory pathway, they do not provide further information about encoding of more temporally complex signals. On the other hand, because the brainstem response to speech provides objective information about how the sound structure of speech syllables is encoded by the auditory system, it can be used to diagnose auditory processing deficits despite normal processing of click stimuli. Thus, brainstem responses to both sounds provide objective and complementary information about sound encoding in the auditory system.

Acknowledgments

Supported by NIH/NIDCD RO1-01510 and the Hugh Knowles Center for Clinical and Basic Science in Hearing and its Disorders, Northwestern University. We would like to thank all members of the Auditory Neuroscience Laboratory for their help in the many stages of the study and the children and families who participated in this study.

References

Ahissar M, Protopapas A, Reid M, Merzenich MM. Auditory processing parallels reading abilities in adults. Proceedings of the National Academy of Sciences of the United States of America 2000; 97: 6832-6837.

Amitay S, Ahissar M, Nelken I. Auditory processing deficits in reading disabled adults. Journal of the Association for Research in Otolaryngology 2002; 3: 302-320.

Baldeweg T, Richardson A, Watkins S, Foale C, Gruzelier J. Impaired auditory frequency discrimination in dyslexia detected with mismatch evoked potentials. Annals of Neurology 1999; 45: 495-503.

Banai K, Ahissar M. Poor frequency discrimination probes dyslexics with particularly impaired working memory. Audiology and Neurotology 2004; 9: 328-340.

Banai K, Nicol T, Zecker SG, Kraus N. Brainstem timing: Implications for cortical processing and literacy. Journal of Neuroscience 2005; 25: 9850-7.

Brandt J, Rosen JJ. Auditory phonemic perception in dyslexia: categorical identification and discrimination of stop consonants. Brain and Language 1980; 9: 324-337.

Brown L, Sherbenou R, Johnsen S. Test of Nonverbal Intelligence Third Edition. Austin, TX: PRO-ED, Inc., 1997.

Cunningham J, Nicol T, Zecker SG, Kraus N. Neurobiologic responses to speech in noise in children with learning problems: deficits and strategies for improvements. Clinical Neurophysiology 2001; 112: 758-767.

Gorga M, Abbas P, Worthington D. Stimulus calibration in ABR measurements. In: Jacobson J, editor. The auditory brainstem response. San Diego, CA: College-Hill Press, 1985: 49-62.

Gorga MP, Kaminski JR, Beauchaine KA, Jesteadt W, Neeley ST. Auditory brainstem responses from children three months to three years of age: II. Normal patterns of response. Journal of Speech and Hearing Research 1989; 50: 346-350.

Grøntved A, Walter B, Gronborg A. Auditory brain stem responses in dyslexic and normal children. A perspective clinical investigation. Scandinavian Audiology 1988a; 17: 53-54.

Grøntved A, Walter B, Gronborg A. Normal ABR's in dyslexic children. Acta Oto-laryngologica Supplementum 1988b; 449: 171-173.

Hall J. Handbook of Auditory Evoked Responses. Needham Heights, MA: Allyn and Bacon, 1992.

Heim S, Keil A. Large-scale neural correlates of developmental dyslexia. European Child & Adolescent Psychiatry 2004; 13: 125-140.

Hood LJ. Clinical applications of the auditory brainstem response. San Diego, CA: Singular Publishing Group, Inc., 1998.

Jacobson J. The Auditory Brainstem Response. San Diego, CA: College-Hill Press, 1985.

Jerger S, Martin RC, Jerger J. Specific auditory perceptual dysfunction in a learning disabled child. Ear and Hearing 1987; 8: 78-86.

Jirsa RE. Maximum length sequences-auditory brainstem responses from children with auditory processing disorders. Journal of the American Academy of Audiology 2001; 12: 155-164.

Johnson KL, Nicol T, Kraus N. The brainstem response to speech: A biological marker of auditory processing. Ear and Hearing 2005; 26: 424-434.

Johnson KL, Nicol TG, Zecker SG, Wright BA, Kraus N. Brainstem timing in learning impaired children with excessive auditory backward masking [Abstract]. Association for Research in Otolaryngology 2004; 27: 118.

Josey A. Auditory brainstem response in site of lesion testing. In: Katz J, editor. Handbook of clinical audiology. Baltimore, MD: Williams and Wilkins, 1985: 534-548.

King C, Warrier CM, Hayes E, Kraus N. Deficits in auditory brainstem pathway encoding of speech sounds in children with learning problems. Neuroscience Letters 2002; 319: 111-115.

Klatt D. Software for cascade/parallel formant synthesizer. Journal of the Acoustic Society of America 1980; 67: 971-975.

Krishnan, A. Human frequency-following responses: representation of steady-state synthetic vowels. Hearing Research 2002; 166; 192-201.

Kraus N, McGee TJ, Carrell TD, Zecker SG, Nicol TG, Koch DB. Auditory neurophysiologic responses and discrimination deficits in children with learning problems. Science 1996; 273: 971-973.

Kraus N, Nicol T. Brainstem origins for cortical 'what' and 'where' pathways in the auditory system. Trends in Neuroscience 2005; 28: 176-181.

Lachmann T, Berti S, Kujala T, Schroger E. Diagnostic subgroups of developmental dyslexia have different deficits in neural processing of tones and phonemes. International Journal of Psychophysiology 2005; 56: 105-120.

Lauter JL, Wood SB. Auditory-brainstem synchronicity in dyslexia measured using the REPs/ABR protocol. Annals of the New York Academy of Sciences 1993; 682: 377-379.

Marler JA, Champlin CA. Sensory processing of backward-masking signals in children with language-learning impairment as assessed with the auditory brainstem response. Journal of Speech, Language and Hearing Research 2005; 48: 189-203.

Mason SM, Mellor DH. Brain-stem, middle latency and late cortical evoked potentials in children with speech and language disorders. Electroencephalography and Clinical Neurophysiology 1984; 59: 297-309.

McAnally KI, Stein JF. Scalp potentials evoked by amplitude-modulated tones in dyslexia. Journal of Speech and Hearing Research 1997; 40: 939-945.

Menell P, McAnally KI, Stein JF. Psychophysical sensitivity and physiological response to amplitude modulation in adult dyslexic listeners. Journal of Speech, Language and Hearing Research 1999; 42: 797-803.

Møller A, Jannetta P. Neural Generators of the Auditory Brainstem Response. In: Jacobson J, editor. The Auditory Brainstem Response. San Diego, CA: College-Hill Press, 1985: 13-32.

Musiek FE. Auditory evoked responses in site-of-lesion assessment. In: Rintelmann WF, editor. Hearing assessment. Austin, TX: PRO-ED, 1991: 383-427.

Nagarajan S, Mahncke H, Salz T, Tallal P, Roberts T, Merzenich M. Cortical auditory signal processing in poor readers. Proceedings of the National Academy of Sciences 1999; 96: 6483-6488.

Paul I, Bott C, Heim S, Eulitz C, Elbert T. Reduced hemispheric asymmetry of the auditory N260m in dyslexia. Neuropsychologia 2005; in press.

Purdy S, Kelly A, Davies M. Auditory brainstem response, middle latency response, and late cortical evoked potentials in children with learning disabilities. Journal of the American Academy of Audiology 2002; 13: 367-382.

Ramus F, Rosen S, Dakin SC, Day BL, Castellote JM, White S, et al. Theories of developmental dyslexia: insights from a multiple case study of dyslexic adults. Brain 2003; 126: 841-865.

Russo N, Nicol T, Musacchia G, Kraus N. Brainstem responses to speech syllables. Clinical Neurophysiology 2004; 115: 2021-2030.

Starr A, Don M. Brainstem potentials evoked by acoustic stimuli. In: Picton TW, editor. Handbook of electroencephalography and clinical neurophysiology. Amsterdam: Elsevier, 1988: 97-150.

Tait CA, Roush J, Johns J. Normal ABR's in children classified as learning disabled. The Journal of Auditory Research 1983; 23: 56-62.

Tallal P. Auditory temporal perception, phonics, and reading disabilities in children. Brain and Language 1980; 9: 182-198.

Torgesen JK. Learning Disabilities: Historical and Conceptual Issues. In: Wong BYL, editor. Learning about Learning Disabilities. San Diego, CA: Academic Press, 1991.

Wagner R, Torgesen J, Rashotte C. Comprehensive Test of Phonological Processing. Austin, TX: PRO-ED, Inc., 1999.

Wible B, Nicol T, Kraus N. Abnormal neural encoding of repeated speech stimuli in noise in children with learning problems. Clinical Neurophysiology 2002; 113: 485-494.

Wible B, Nicol T, Kraus N. Atypical brainstem representation of onset and formant structure of speech sounds in children with language-based learning problems. Biological Psychology 2004; 67: 299-317.

Wible B, Nicol T, Kraus N. Correlation between brainstem and cortical auditory processes in normal and language-impaired children. Brain 2005; 128: 417-423.

Wilkinson G. Wide Range Achievement Test-3. Wilmington, DE: Jastak, 1993.

Woodcock R, Johnson M. Woodcock-Johnson Psycho-Educational Battery-Revised. Allen, TX: DLM Teach Resources, 1989.

Wright BA, Lombardino LJ, King WM, Puranik CS, Leonard CM, Merzenich MM. Deficits in auditory temporal and spectral resolution in language-impaired children. Nature 1997; 387: 176-178.

AUDITORY TRAINING IMPROVES NEURAL TIMING IN THE HUMAN BRAINSTEM

Published in *Behavioural Brain Research*

Nicole M. Russo[a,b,c,f], Trent G. Nicol[a,c], Steven G. Zecker[c], Erin A. Hayes[a,b,c], Nina Kraus[a-e]

[a]Auditory Neuroscience Laboratory, [b]Institute for Neuroscience, [c]Departments of Communication Sciences, [d]Neurobiology and Physiology and [e]Otolaryngology, Northwestern University

Abstract

The auditory brainstem response reflects neural encoding of the acoustic characteristic of a speech syllable with remarkable precision. Some children with learning impairments demonstrate abnormalities in this preconscious measure of neural encoding especially in background noise.

This study investigated whether auditory training targeted to remediate perceptually-based learning problems would alter the neural brainstem encoding of the acoustic sound structure of speech in such children. Nine subjects, clinically diagnosed with a language-based learning problem (e.g., dyslexia), worked with auditory perceptual training software. Prior to beginning and within three months after completing the training program, brainstem responses to the syllable /da/ were recorded in quiet and background noise. Subjects underwent additional auditory neurophysiological, perceptual, and cognitive testing. Ten control subjects, who did not

participate in any remediation program, underwent the same battery of tests at time intervals equivalent to the trained subjects.

Transient and sustained (frequency-following response) components of the brainstem response were evaluated. The primary pathway afferent volley – neural events occurring earlier than 11 ms after stimulus onset – did not demonstrate plasticity. However, quiet-to-noise inter-response correlations of the sustained response (~11–50 ms) increased significantly in the trained children, reflecting improved stimulus encoding precision, whereas control subjects did not exhibit this change. Thus, auditory training can alter the preconscious neural encoding of complex sounds by improving neural synchrony in the auditory brainstem. Additionally, several measures of brainstem response timing were related to changes in cortical physiology, as well as perceptual, academic, and cognitive measures from pre-to post-training.

Keywords: Auditory brainstem response; Neural timing; Plasticity; Speech; Auditory training; Frequency-following response; Reading disability

Introduction

This study addresses several questions: Is there plasticity in the neural encoding of sound in the human auditory brainstem? If so, is this manifested in a way that can be readily measured? Can the brainstem representation of speech-sound structure in children with learning disabilities be altered by work with a commercially available auditory training regimen?

Auditory training has been shown to alter the neural encoding of sound structure at the cortical level. Cortical plasticity has been established in both animals [4,7,11,15,22,31,36,42,43] and humans [25,39]. Cortical changes have accompanied perceptual learning of nonnative speech sounds in adults [59] and improved auditory perception in children with learning problems [25,61]. However, neural plasticity is not necessarily restricted to the cortex. The auditory cortex receives sensory input via the thalamocortical loop and there is a precedent for subcortical plasticity from a number of animal studies [3,10,12–14,21,24,27–30,35,45–47,56,62]. In general, it is thought that subcortical plasticity is short-term. However, once conditioned, the association of a sound with a meaning causes long-term cortical changes. Furthermore, it has been suggested that there is an interaction between ascending auditory pathways and the descending corticofugal system, as well as interactions with the amygdala and basal forebrain [56]. A positive feedback loop involving lateral inhibition modulates subcortical and cortical activity. The extent to which plasticity at subcortical regions directly influences the cortex or vice versa has yet to be determined. Whatever the mechanism, current research supports a relationship between cortical and subcortical plasticity.

Subcortical plasticity in the medial geniculate body (MGB), which synapses directly onto auditory cortex, occurred with classical conditioning in rats [10]. These experience-dependent changes persisted for 45 days. Later, using a guinea pig model, Edeline and Weinberger extensively investigated plasticity in the dorsal (MGd), ventral (MGv), and medial (MGm) divisions of the MGB in response to associative cardiac conditioning to specific frequencies. Each area of the MGB experienced changes in receptive field properties after only a short

conditioning period. Changes in the nonprimary pathway (MGd) were resilient and persisted at the one-hour post-test session [12], while changes in the primary pathway (MGv) were susceptible to decay after 1 h [13]. Both short-and long-term changes were seen in the MGm, reflecting the broad-and fine-tuned bandwidth variation of cells in this area [14]. Edeline and Weinberger concluded that the subdivisions of the MGB act in conjunction with each other and that the significance of the stimulus affects the duration of the change.

Plasticity in the cochlear nucleus has been demonstrated in a decerebrate preparation using a cat as the animal model [3,27,28]. This basic paradigm resulted in the expression of habituation and spontaneous recovery in the cells of the cochlear nucleus in response to repetitive stimulation. Alterations in neural connectivity following cochlear ablation demonstrate plasticity in even lower subcortical structures. Unilateral cochlear removal in ferrets produced changes in the number of neurons projecting to the contralateral inferior colliculus [46,47]. Ablation of the cochlear nucleus in rats resulted in new patterns of synaptic connections within the brainstem [29,30]. Illing et al. [30] further explored brainstem plasticity after cochlear ablation in rats. They observed plasticity in the superior olivary complex, ventral and dorsal cochlear nucleus, and inferior colliculus via the increased presence of GAP-43, which is abundant during synaptogenesis both in development and remodeling.

Plasticity at the level of the inferior colliculus has been observed in barn owls. Behavioral changes in sound localization following filtering or ear occlusion were accompanied by changes

in auditory space maps within the inferior colliculus [21,35,45].

These animal studies showed changes occurring in receptive fields, space maps, and synaptic activity and connectivity at the first levels of sensory processing. In the only human study to our knowledge, Khalfa et al. investigated the modulation of auditory periphery by higher cortical regions in epileptic patients following resection surgery [32]. They were able to demonstrate reciprocal relationships between changes in the medial olivocochlear system and auditory cortex. Using transiently evoked otoacoustic emission recordings and equivalent attenuation calculations, they were able to assess effects of the surgery on the medial olivocochlear nucleus. Specifically, they showed evidence for corticofugal influence on the medial olivocochlear nucleus and associated changes in speech processing both in quiet and noise.

Thus, considerable evidence suggests that neuronal activity occurring in the human auditory midbrain may be dynamic. The current study was designed to investigate plasticity in the physically intact auditory system by capitalizing on the ability to quantify temporal changes using evoked potentials. A strength of evoked potentials is their use in quantifying neural synchrony and timing in the encoding of complex stimuli, such as speech.

Specific aspects of the sound structure are maintained and reflected in the neural code of the auditory brainstem [51]. The brainstem response to a speech sound consists of two components, the onset and the frequency-following response (FFR), which represent transient and sustained processes, respectively. Transient responses, with precision on the order of tenths of milliseconds,

represent primarily the response to discrete events in the stimulus, such as the stimulus onset and the successive modulations caused by the vibration of the vocal folds. Sustained response components last for the duration of a periodic stimulus and reflect the overall integrity of the response with respect to the stimulus.

A speech syllable can be divided into transient and sustained portions – consonants and vowels – that share some characteristics with the brainstem response components. Consonants are rapid, transient, and generally aperiodic features of speech; they are represented by the transient components of the brainstem response and are easily disrupted by noise. Vowels are periodic, sustained signals; they are represented by the sustained features of the brainstem response, are generally much larger in amplitude than consonants and are more resistant to noise.

Stop consonants are difficult to perceive, especially for people with learning disabilities [2,6,49]. Children with language-based learning problems often exhibit deficits in auditory perception and the neural encoding of speech sounds at both cortical and brainstem levels [5,20,38,52,57,64,68], especially when background noise is introduced [1,33,63]. Commercial auditory training programs have been developed to provide remediation for auditory perception and related learning deficits [8,18,19,44,57]. The physiological consequence of this kind of training is little understood. Thus, testing children before and after undergoing such training offers an ideal opportunity to examine neural plasticity at the level of the auditory brainstem.

Normative values and test-retest reliability for the brainstem measures in this study have been established and provide a means for determining the degree to which brainstem responses may be expected to change over time [33,51]; pre- to post-training changes that exceed test–retest changes can be attributed to auditory training. Moreover, because not all children benefit in the same way from training programs, it is important to determine what pre-training neurophysiological measures are markers for successful training.

Speech-evoked brainstem activity was obtained before and after children with learning disabilities participated in a commercial auditory training regimen. Both transient and sustained components of the brainstem response to the syllable /da/ presented in quiet and in background noise were assessed. Relationships of brainstem measures to improvements in cognitive, perceptual, and academic achievement tests were also explored.

Materials and Methods

Subjects and training regimen

Nineteen children, 8–12 years old, were included in this study. All of the subjects were native English speakers, with normal IQ (≥85 on Brief Cognitive scale or Test of Nonverbal Intelligence; range 85–135), and had normal hearing thresholds at or below 20 dB HL for octaves from 500 to 4000 Hz. The experimental group comprised nine children with learning disabilities (LD) based on diagnoses by outside professionals (child psychologists, neurologists, etc.) and their performance on study-internal measures of learning and academic achievement

(see Section 2.3). Consent and assent were obtained from the parent(s) or legal guardian(s) and the children. The Institutional Review Board of Northwestern University approved all research.

Children in the experimental group participated in 35–40 independently supervised one-hour sessions of Earobics [9] over an 8-week period. Earobics is a commercial auditory training program that provides training through interactive computer games of phonological awareness, auditory processing, and language processing skills. The stimuli are presented in quiet and background noise, with both visual and auditory feedback. Children listen to sounds while playing interactive, animated computer games; they match sounds (indicating alike or different) by clicking the computer mouse on appropriate pictures or sound representations they hear. Earobics is a two-step program; Step 1 has six interactive games covering phonological awareness and processing, while five games comprise Step 2, which further develops the skills trained in Step 1 and concentrates more on language processing skills to help individuals better interpret spoken and written language [9]. Children in the experimental group underwent auditory neurophysiological and perceptual/cognitive testing prior to and within three months following completion of the training program.

The control group underwent the same battery of tests at a time interval equivalent to the trained subjects, but did not participate in the training program. This group (n = 10) consisted of both normal learning (NL; n = 5) and LD (n =5) children who met the same inclusion criteria as those in the experimental group.

Neurophysiological testing

Auditory brainstem and cortical evoked potentials were evaluated in response to the speech syllable /da/ presented in quiet and background noise.

Auditory brainstem response

The brainstem response was elicited by the synthesized [34] speech stimulus /da/ (Fig. 1, top). The stimulus duration was 40 ms. Randomly alternating polarities were presented (Neuroscan, Stim, Compumedics) to the right ear through an insert earphone (ER-3, Etymotic Research) at 80 dB SPL with a 51 ms inter-stimulus interval. The syllable was presented in two conditions, quiet and with white Gaussian background noise (+5 dB SNR). The response was differentially recorded from Cz-to-ipsilateral earlobe, with the forehead as ground. Three thousand sweeps per polarity were collected (Neuroscan, Scan, Compumedics) in each noise condition. The sampling rate was 20,000 Hz and responses were online filtered from 100 to 2000 Hz. Trials with activity greater than 35 □V were online rejected. Responses to alternating polarity stimuli were added together to create a mainly neural response [23]. Throughout the testing session, the children watched a video of their choice and listened to the soundtrack at less than 40 dB SPL in the non-test ear.

Transient response. The brainstem response to /da/ consists of six major transient peaks (A–F) following the familiar I–V series. These peaks represent synchronized neural activity in response to the phonetic/acoustic characteristics of the speech syllable and represent peaks within the stimulus with remarkable precision. Peaks V, A, C, and F are the most reliable peaks in the

response, exhibiting small latency variability and excellent detectability in all subjects [33]. These peaks (Fig. 1) were evaluated both in terms of timing (latency) and magnitude (amplitude). The VA complex was further analyzed by interpeak latency, area, amplitude, and slope. A wavelet-denoising technique derived from Qian Quiroga and Garcia [48] was used to aid in determining peak latencies and amplitudes of responses elicited in noise.

Sustained response. The sustained FFR component of the response (11.5–46.5 ms) (Fig. 1) was evaluated both by magnitude and timing measures. Magnitude was evaluated in two ways. RMS amplitude was calculated over the FFR epoch. The amplitude of the spectral component encompassing the fundamental frequency of the stimulus (F0 = 103–121 Hz) was measured by fast Fourier transformation analysis. Timing also was assessed in two ways, using a cross-correlation technique. Stimulus-to-response correlations were measured, using the 10–40 ms portion of the stimulus, and the highest correlation achieved within a response lag of 6–9 ms was obtained. Quiet-to-noise inter-response correlations were also analyzed over a response range of 10–40 ms, with a noise response lag of up to 2 ms. Specific details of the methods and normative values are discussed elsewhere [5,33,51].

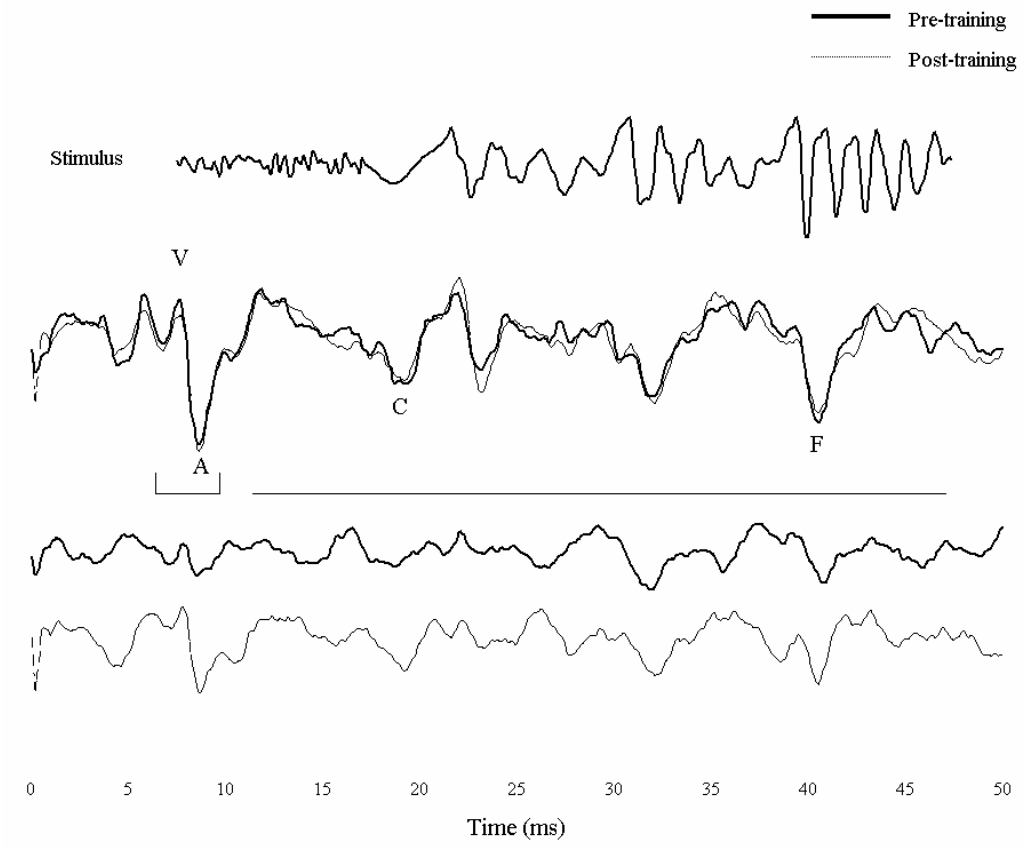

Fig. 1. Stimulus waveform and grand averages of those subjects whose quiet-to-noise inter-response correlations improved. The stimulus has been shifted to align peaks present in the stimulus with their corresponding response waveform peaks; this shift accounts for a time delay introduced by the amount of time required for the sound to traverse the ear canal to the brainstem. Peaks are labeled, the onset response is bracketed and the FFR is underlined in the quiet response. Waveforms show that the improvement of correlations can be attributed to more accurate encoding of the signal in noise, rather than a change in quiet.

Analysis of plasticity

Plasticity in physiological measures in trained subjects was defined as changes in the neurophysiological response that exceeded those observed in the untrained control subjects. Differences between groups were measured using a repeated measures analysis of variance with test session as the within-subject factor and training group as the between-subject factor. Post-hoc tests were done to establish in which group the significant changes occurred. A criterion of $P < 0.05$ was used. For all statistical analyses involving Pearson correlations, Fisher's transformation was used to convert r-values to z-scores.

Cortical response

Cortical responses to the speech stimulus /da/ presented at 80 dB SPL in quiet and noise (0 dB SNR) were recorded. The interstimulus interval was 590 ms. The sampling rate was 2000 Hz and responses were online filtered from 0.05 to 100 Hz. Cortical activity was recorded from Cz, with a nasal reference and the forehead as ground. Eyeblink was monitored with bipolar supraorbital-to-lateral canthus electrodes. P2N2 amplitudes, latencies, and quiet-to-noise inter-response correlations were measured. Similar to the technique for analyzing the inter-response correlations for the brainstem response, the cortical response to the sound presented in quiet was cross-correlated with the response recorded in background noise. The correlation was calculated over the 100–350 ms range and the highest correlation value achieved within a 50 ms lag was obtained [25]. Spearman correlations were used to identify relationships between brainstem and cortical response measures.

Perceptual and cognitive abilities testing

At both the test and retest sessions, subjects underwent a series of tests that quantified their perceptual and cognitive abilities. Subjects were evaluated on measures of auditory processing (Incomplete Words, Memory for Words, Sound Blending, Listening Comprehension [67]), mental abilities (Brief Cognitive Scale [66]), and academic achievement (Word Attack [67], Reading and Spelling [65]). Other measures of auditory perception included speech discrimination in quiet and in background noise (just-noticeable difference scores along a synthesized /da-ga/ continuum differing in F3 onset frequency, as determined by Parameter Estimation Sequence Tracking [58]), speech identification (perception of Sentences in Noise [1]), and temporal resolution (Backward Masking). These measures have been described in detail elsewhere [1,38,68].

Spearman correlations were used to identify relationships between the brainstem response and cognitive and perceptual measures. For these analyses, if a subject showed a decrease on a perceptual/cognitive score upon retest, their "improvement" was coded as zero to diminish the impact of outliers.

Results

Stability of brainstem measures over time: control group

Test–retest data were collected from control subjects who did not undergo auditory training. NL and LD controls were combined into one control group because the degree of change of test–retest measures was equivalent between the NL and LD controls. Two-tailed, paired t-tests were conducted to establish changes that could be expected to occur over a 3–6-month time interval.

These comparisons revealed that most /da/-elicited brainstem measures are stable over time. No significant differences were found in brainstem measures obtained in quiet, with the exceptions of VA interpeak amplitude and slope (both, $P < 0.02$). Onset response amplitude is known to be variable [53], so this was not a compelling change. Onset amplitudes were thus omitted from analysis of training effects. In background noise, onset responses are often attenuated to a great extent and sometimes eliminated [51]. Therefore, these responses were not evaluated for effects of training. Peaks C and F, however, remained robust and were resistant to background noise. All FFR measures and transient response peaks C and F remained stable in quiet and background noise over the test–retest time interval.

Effects of training on brainstem measures: experimental group

Measures of onset response timing did not change in the experimental group. There was no evidence of training associated changes in responses occurring earlier than 11 ms.

Quiet-to-noise inter-response correlations of the FFR increased significantly for the experimental group after training, but not for the control group (RMANOVA interaction, $F_{approx}(1,17) = 6.67$, $P < 0.02$; post-hoc one-tailed paired t-test, $P < 0.02$ and $P > 0.25$, respectively). Specifically, seven of the nine trained subjects showed this increase (Fig. 2). Increased quiet-to-noise inter-response correlations indicate that timing characteristics of the stimulus became encoded more precisely after training.

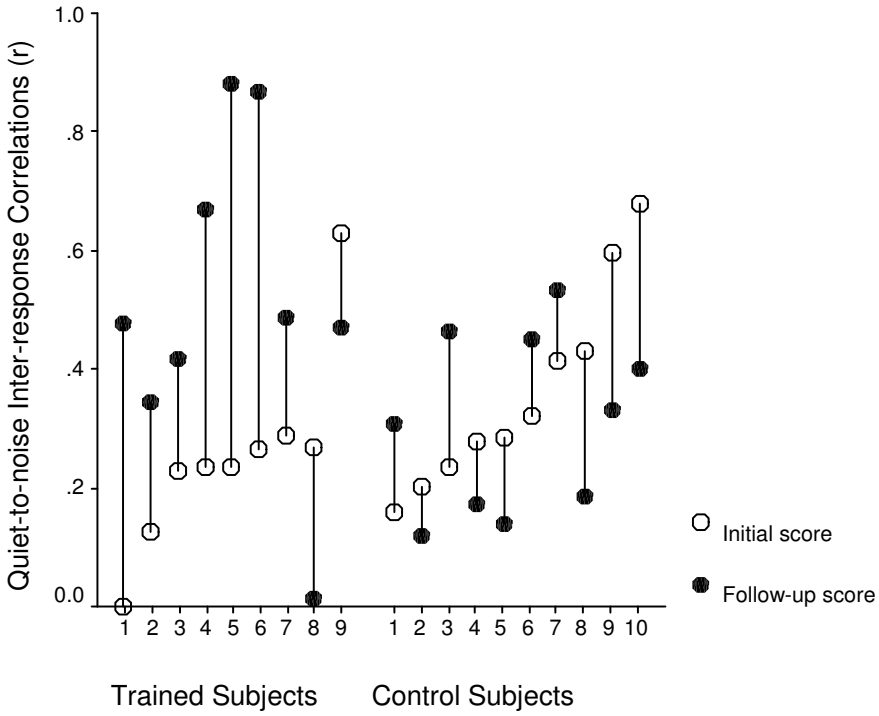

Fig. 2. Improved neural timing in noise. Quiet-to-noise inter-response correlations of trained (left) and control subjects (right). Subcortical changes in the brainstem response occurred in trained subjects, as evidenced by 7 of the 9 subjects with increased inter-response correlations, while control subjects did not change (z′ score conversion; RMANOVA interaction, F_{approx} (1,17)=6.67, $p<0.02$; post hoc one-tailed paired t-test, $p<0.02$ and $p>0.25$, respectively). Increased correlations are indicative of more similarity between quiet and noise responses, suggesting improved encoding in noise.

In order to discern whether an improvement in either the response in quiet or noise contributed more to the overall improvement in the quiet-to-noise inter-response correlations, grand average waveforms were compared (Fig. 1). Visual inspection of these waveforms suggests that

responses in quiet were stable while clearer definition of noise response components emerged following training.

To quantify this observation, partial correlation analyses, controlling for values of pre-training stimulus-to-response correlations in quiet, were performed. A strong and significant relationship was found between inter-response and stimulus-to-response in noise correlations (partial correlation, 0.55, one-tailed $P < 0.01$). However, the stimulus-to-response in quiet showed no such relationship with the inter-response correlations. This confirmed the presumption that an improvement of neural timing in noise made the greater contribution to the overall increase in inter-response correlations. Following training, the overall morphology of the waveform for the response in noise more closely resembled the response in quiet, and thus the stimulus. Additionally, trained subjects showed better wave C peak definition and a later latency in noise, unlike control subjects (RMANOVA interaction, $F_{approx}(1,17)=7.24$, $P < 0.02$; post-hoc one-tailed paired t-test, $P < 0.01$ and $P > 0.25$, respectively). These changes may have contributed to the improved correlation between the quiet and noise responses. The training-associated change in peak C latency in noise was likely a consequence of post-training sharpening of the wave. As can be seen in the pre-training grand averaged waveform (Fig. 1), the region around peak C (approx. 19 ms) was very broad; the peak latency was not clearly identified. Furthermore, the standard deviation of peak C latency in noise decreased post-training. Thus, after training, as the peak became more pronounced, the judgment of its latency became more precise.

Relationships between subcortical and cortical measures

A study conducted by Hayes et al. [25] showed changes in response timing and magnitude in noise following auditory training. This was manifested by increases in cortical quiet-to-noise inter-response correlations and P2N2 amplitudes in noise. Relationships between improvements at both the brainstem and cortical levels were explored. Increases in brainstem quiet-to-noise inter-response correlations were significantly associated with cortical P2N2 amplitude increases in noise (Spearman's $\rho = 0.70$, one-tailed $P < 0.03$). Increases in stimulus-to-brainstem response in noise correlations were associated with increases in cortical amplitudes (Spearman's $\rho = 0.67$, one-tailed $P < 0.03$). Overall, improved subcortical timing was associated with improvements in the cortical response.

Relationships between brainstem responses and behavior

Relationships between training-related brainstem response changes and changes in perceptual and cognitive measures were examined. Additionally, pre-training brainstem response indicators of behavioral improvement were sought.

Children in the trained group demonstrated significant gains on the Incomplete Words, Auditory Processing, and Sentences-in-Noise tests. Although brainstem changes were not directly related to improvements on those particular tests, gains in Listening Comprehension were related to changes in the brainstem response in noise. Decreases in the RMS amplitude of the FFR in noise were accompanied by improved Listening Comprehension scores (Spearman's $\rho = -0.88$, $P < 0.002$), a measure of auditory processing. No other significant relationships were found between brainstem response changes and changes on the other tests of perceptual and cognitive abilities.

Brainstem response markers of training success

Particular pre-training brainstem measures marked children who demonstrated significant training-associated gains in auditory processing and speech discrimination in noise. Children with later peak F latencies in noise demonstrated improvements on Incomplete Words (Spearman's $\rho = -0.90$, $P < 0.001$), while children with larger peak F amplitudes in noise showed improvements in /da-ga/ discrimination in noise (Spearman's $\rho = -0.84$, $P < 0.005$).

Discussion

Measures of transient onset response timing were stable over time and resistant to the effects of training; they did not change in either experimental or control subjects. Auditory training did alter sustained response timing. Brainstem response quiet-to-noise inter-response correlations, as well as FFR peak C latency in noise, differed between test sessions in children who received training, but not in control subjects. Training did not alter sustained response magnitudes.

Improved neural timing in noise

Auditory training appears to alter the brainstem response to speech sounds. Specifically, neural encoding became more resistant to the deleterious effects of background noise following training. Increases in quiet-to-noise inter-response correlations represent greater timing precision in the FFR in noise after training.

Certain assumptions can be made about the nature of plasticity within the auditory brainstem based on the latency ranges over which changes did and did not occur following auditory perceptual training. Onset responses to /da/, occurring within the first 11 ms post-stimulus onset, were relatively stable over time and were also unaffected by training. Thus, a response that arises exclusively from the primary afferent volley did not demonstrate plasticity; neural events occurring so early in the processing of auditory stimulation may be hardwired. Unlike the onset response, the FFR element of the brainstem response was found to be dynamic. Auditory training altered the neural encoding of the harmonic, periodic aspects of sound occurring 12–40 ms post-stimulus onset.

Where do the changes occur?

Isolating the precise source of neural plasticity in the auditory brainstem cannot be accomplished with far-field recordings, although the time frame of the plasticity provides considerable information regarding the likely neuroanatomical contributions. Because no changes occurred earlier than the first 12 ms post-stimulus onset, it is plausible that the inferior colliculus itself is the locus of plastic activity [37]. However, it is also possible that plasticity at sites peripheral to the inferior colliculus may be contributing to the plasticity shown in this study. Some studies have shown effects of attention on cochlear activity [17,40,41]. Galbraith et al. [16] suggests that such short latency attentional effects may affect the FFR component of the brainstem response to vowels. Therefore modulation of cochlear hair cells might influence early processing within the superior olivary complex and thus alter the activity of the inferior colliculus. Hoormann et al. [26] also corroborates the concept of early attentional modulation of the FFR.

Cortical feedback can also induce plasticity within the FFR. Intracranial recordings in human auditory cortex have observed activation as early as 12 ms in response to clicks and tone bursts. Steinschneider et al. [54] reported a similar time frame in response to a /da/ syllable. Given that the initial response in cortex occurs at such a short latency, it can be theorized that cortical feedback may regulate neural activity as early as the timeframe seen in the present study (e.g., within the first 30 ms). While the cortex may modulate activity, the locus of plasticity is not likely rostral to the inferior colliculus since the MGB and auditory cortex do not phase-lock at rates as fast as fundamental and first formant frequencies [50,60].

The corticofugal descending system is critical in manipulating signal encoding via positive feedback or lateral inhibition mechanisms [55]. Once trained or conditioned, egocentric selection [69,70] allows for the cortex to recognize the behavioral significance of an acoustic stimulus and then fine-tune its own input by altering the sound representation at lower levels. Specifically, the cortex modulates subcortical areas that encode basic stimulus features and thus improves subsequent cortical representation. Even a short-term subcortical change, lasting 1–3 h, is sufficient to influence long-term cortical changes [13,14,55,70]. Although it is still unknown precisely how corticofugal modulation is initiated, the evidence remains that subcortical regions are malleable with training and that modulation may occur in multiple domains (frequency, time, etc.).

To our knowledge training-associated neural plasticity at the level of the brainstem in humans has not been previously identified. However, extensive animal research, as reviewed above, has demonstrated regions of plasticity at subcortical levels. Classical conditioning, auditory deprivation, and cochlear ablation studies support the idea that plasticity does occur subcortically and may affect cortical processing directly. Alternatively, cortical and subcortical activity may modulate each other through corticofugal loops. The aforementioned animal studies together with the present work demonstrate plasticity in the auditory brainstem and support the notion that early sensory processing is malleable.

Behavioral ramifications

The relationships between brainstem changes and behavioral measures support the idea that pre-conscious alteration of the brainstem response affects auditory perception. Gains in Listening Comprehension were related to a reduction in the sustained response RMS amplitude in noise. During the prestimulus period, RMS amplitudes did not change between test sessions (paired t-test, $P = 0.34$), indicating that this reduction was confined to stimulus-evoked activity and not an overall reduction in physiological noise due to factors such as subject state or electrode impedance. Clearly, noisy listening environments impair perception. Subjects without extraneous noise in their brainstem response, as suggested by lower RMS amplitudes and sharper peak definition, were able to more accurately decipher what they heard, as evidenced by improved Listening Comprehension scores. A more precise brainstem response in noise may benefit the listener by providing a more accurate representation of the acoustic characteristics of the stimulus. This study also suggested that particular pre-training brainstem response measures in

both quiet and noise may be related to improvements in measures of auditory processing and speech discrimination.

Clinicians and parents might be able to streamline their children's training programs based on information gained from pre-training speech-evoked brainstem response screening. This study and other related work from our laboratory [25,33] indicate that children with delayed brainstem timing are particularly likely to profit from auditory training. Thus, brainstem response screening may serve as a means to identify children for auditory training rehabilitation. Eventually, one might envision designing a training regimen tailored to a child's particular needs.

Not all children who went through auditory training demonstrated neurophysiological changes at the level of the brainstem. The amount of time between finishing training and returning for neurophysiological testing did not affect the outcome. The two subjects who did not show improved neural timing were in the middle of the group with respect to test–retest interval. Thus, the elapsed time appeared not to influence whether or not the subject exhibited timing improvements in the brainstem response. Because behavioral improvements could occur in the absence of neurophysiological changes, these changes may be sufficient, but are not entirely necessary for behavioral gains. However, a considerably larger population needs to be assessed before the "sufficient versus necessary" question can be answered definitively. It is possible that those children who showed no changes in brainstem activity had deficits that were not addressed by the training they received. Alternatively, those subjects' learning and auditory perception problems may not have stemmed from an auditory encoding deficiency at the brainstem level.

Future longitudinal investigations may determine whether longer training sessions (hours per day or number of weeks) or repeated training sessions spread out over multiple 8-week periods would, in fact, alter the brainstem responses in the children who did not show physiological changes in this study. Extended research may fill in the gaps pertaining to the rigidity of the onset of the brainstem response to training. Follow-up testing would offer further information about the resilience of the neurophysiological and corresponding perceptual and behavioral effects of training.

Pre-conscious modification of sensory processing, prior to cognitive processing, may help overcome higher level deficits. Previous research [25] showed that children who went through Earobics training experienced changes in cortical responses to speech syllables, including accelerated maturation of the response, larger amplitudes, and improved quiet-to-noise inter-response correlations. The relationship between subcortical and cortical improvements reported here suggest that alterations in the brainstem response could have contributed to a more intact neural representation of sound at the cortical level.

Extensions of this work

This work demonstrated the existence of plasticity at the level of the human auditory brainstem and that auditory training can improve neural timing in response to sounds. There are broad-reaching implications. Previous work has shown that specific measures of the brainstem response can serve as biological markers that can identify a subset of language-impaired children with encoding deficits [6,33,64]. Consequently, the brainstem response to speech could be used in

early detection of children “at risk” for these learning problems and who may benefit from auditory training. Thus, remediation can begin before children reach school age. Regardless of the age of identification and remediation, any changes in the brainstem response may be used as an objective monitor of auditory training success.

Although the children in this study underwent a general mode of auditory training, effects were transferable among sounds, since it was associated with alteration of the response to the laboratory test syllable /da/. Even so, one can imagine greater success of training programs that target specific difficulties or encoding deficits. For example, training via cue enhancement is used in other auditory training programs. The brainstem response employed here could be informative regarding effects of different forms of auditory training. Moreover, auditory training could be targeted at enhancing specific acoustic characteristics that are not encoded accurately at the brainstem. Finally, this experimental approach can be applied to other populations in which perceptual learning relevant to language and communication are of interest (e.g., second-language learning, aging, cochlear implant recipients, autistic individuals, etc.).

Conclusions

Neural encoding of sound in the human brainstem appears to be modified by auditory training. This study used measures of timing and magnitude of the brainstem response to identify possible mechanisms of brainstem plasticity. In addition, measures of brainstem plasticity were discovered to be associated with perceptual and cognitive changes. The conclusions drawn from this data set complement results drawn from cortical and subcortical animal and human studies

that indicate learning-associated plasticity in the auditory pathway. Moreover, this study provides evidence that commercially available auditory training can alter the preconscious neural encoding of sound by improving neural synchrony in the human auditory brainstem. The National Institute of Health NIDCD R01-01510 supported this research.

Acknowledgements

We would like to thank the children who participated in this study and their families. We would also like to acknowledge Brad Wible, Pam Horstmann, and Erika Skoe who assisted with the data collection and processing.

References

[1] Bradlow AR, Kraus N, Hayes E. Speaking clearly for children with learning disabilities: sentence perception in noise. J Speech Lang Hear Res 2003;46(1):80–97.

[2] Bradlow AR, Kraus N, Nicol TG, McGee TJ, Cunningham J, Zecker SG, et al. Effects of lengthened formant transition duration on discrimination and neural representation of synthetic CV syllables by normal and learning-disabled children. J Acoust Soc Am 1999;106(4(Pt 1)):2086–96.

[3] Buchwald JS, Humphrey GL. Response plasticity in cochlear nucleus of decerebrate cats during acoustic habituation procedures. J Neurophysiol 1972;35(6):864–78.

[4] Cansino S, Williamson SJ. Neuromagnetic fieldsreveal cortical plasticity when learning an auditory discrimination task. Brain Res 1997;764(1–2):53–66.

[5] Cunningham J, Nicol T, Zecker S, Kraus N. Speech-evoked neurophysiologic responses in children with learning problems: development and behavioral correlates of perception. Ear Hear 2000;21(6):554–68.

[6] Cunningham J, Nicol T, Zecker SG, Bradlow A, Kraus N. Neurobiologic responses to speech in noise in children with learning problems: deficits and strategies for improvement. Clin Neurophysiol 2001;112(5):758–67.

[7] Diamond DM, Weinberger NM. Role of context in the expression of learning-induced plasticity of single neurons in auditory cortex. Behav Neurosci 1989;103(3):471–94.

[8] Diehl S. Listen learn? A software review of Earobics. Lang Speech Hear Serv Schools 1999;30:108–16.

[9] Earobics. http://www.earobics.com/. Evanston, IL: Cognitive Concepts.

[10] Edeline JM, Neuenschwander-el Massioui N, Dutrieux G. Discriminative long-term retention of rapidly induced multiunit changes in the hippocampus, medial geniculate and auditory cortex. Behav Brain Res 1990;39(2):145–55.

[11] Edeline JM, Pham P, Weinberger NM. Rapid development of learning-induced receptive field plasticity in the auditory cortex. Behav Neurosci 1993;107(4):539–51.

[12] Edeline JM, Weinberger NM. Subcortical adaptive filtering in the auditory system: associative receptive field plasticity in the dorsal medial geniculate body. Behav Neurosci 1991;105(1):154–75.

[13] Edeline JM, Weinberger NM. Thalamic short-term plasticity in the auditory system: associative returning of receptive fields in the ventral medial geniculate body. Behav Neurosci 1991;105(5):618–39.

[14] Edeline JM, Weinberger NM. Associative retuning in the thalamic source of input to the amygdala and auditory cortex: receptive field plasticity in the medial division of the medial geniculate body. Behav Neurosci 1992;106(1):81–105.

[15] Edeline JM, Weinberger NM. Receptive field plasticity in the auditory cortex during frequency discrimination training: selective retuning independent of task difficulty. Behav Neurosci 1993;107(1):82–103.

[16] Galbraith GC, Bhuta SM, Choate AK, Kitahara JM, Mullen Jr TA. Brain stem frequency-following response to dichotic vowels during attention. NeuroReport 1998;9(8):1889–93.

[17] Giard MH, Collet L, Bouchet P, Pernier J. Auditory selective attention in the human cochlea. Brain Res 1994;633(1–2):353–6.

[18] Gillam RB, Crofford JA, Gale MA, Hoffman LM. Language change following computer-assisted language instruction with Fast ForWord or Laureate Learning Systems software. Am J Speech Lang Pathol 2001;10(3):231–47.

[19] Gillam RB, Loeb DF, Friel-Patti S. Looking back: a summary of five exploratory studies of Fast ForWord. Am J Speech Lang Pathol 2001;10(3):269–73.

[20] Godfrey JJ, Syrdal-Lasky AK, Millay KK, Knox CM. Performance of dyslexic children on speech perception tests. J Exp Child Psychol 1981;32(3):401–24.

[21] Gold JI, Knudsen EI. A site of auditory experience-dependent plasticity in the neural representation of auditory space in the barn owl's inferior colliculus. J Neurosci 2000;20(9):3469–86.

[22] Gonzalez-Lima F, Scheich H. Neural substrates for tone-conditioned bradycardia demonstrated with 2-deoxyglucose. II. Auditory cortex plasticity. Behav Brain Res 1986;20(3):281–93.

[23] Gorga MP, Abbas PJ, Worthington DW. Stimulus calibrations in ABR measurements. In: Jacobson JT, editor. The Auditory Brainstem Response. San Diego: College-Hill Press; 1985. p. 49–62.

[24] Halas ES, Beardsley JV, Sandlie ME. Conditioned neuronal responses at various levels in conditioning paradigms. Electroencephalogr Clin Neurophysiol 1970;28(5):468–77.

[25] Hayes EA, Warrier CM, Nicol TG, Zecker SG, Kraus N. Neural plasticity following auditory training in children with learning problems. Clin Neurophysiol 2003;114(4):673–84.

[26] Hoormann J, Falkenstein M, Hohnsbein J. Effect of selective attention on the latency of human frequency-following potentials. NeuroReport 1994;5(13):1609–12.

[27] Huang CM, Buchwald JS. C comparison of acoustic nerve and cochlear nucleus responses during acoustic habituation. Brain Res 1979;173(3):562–6.

[28] Huang CM, Buchwald JS. Changes of acoustic nerve and cochlear nucleus evoked potentials due to repetitive stimulation. Electroencephalogr Clin Neurophysiol 1980;49(1–2):15–22.

[29] Illing RB, Forster CR, Horvath M. Evaluating the plasticity potential of the auditory brain stem nucleus in the rat. Am J Otol 1997;18(Suppl. 6):52–3.

[30] Illing RB, Horvath M, Laszig R. Plasticity of the auditory brainstem: effects of cochlear ablation on GAP-43 immunoreactivity in the rat. J Comp Neurol 1997;382(1):116–38.

[31] Irvine DR, Rajan R, Brown M. Injury-and use-related plasticity in adult auditory cortex. Audiol Neurootol 2001;6(4):192–5.

[32] Khalfa S, Bougeard R, Morand N, Veuillet E, Isnard J, Guenot M, et al. Evidence of peripheral auditory activity modulation by the auditory cortex in humans. Neuroscience 2001;104(2):347–58.

[33] King C, Warrier CM, Hayes E, Kraus N. Deficits in auditory brainstem pathway encoding of speech sounds in children with learning problems. Neurosci Lett 2002;319(2):111–5.

[34] Klatt D. Software for cascade/parallel formant synthesizer. J Acoust Soc Am 1980:67971–5.

[35] Knudsen EI, Zheng W, DeBello WM. Traces of learning in the auditory localization pathway. Proc Natl Acad Sci USA 2000;97(22):11815–20.

[36] Kraus N, Disterhoft JF. Response plasticity of single neurons in rabbit auditory association cortex during tone-signalled learning. Brain Res 1982;246(2):205–15.

[37] Kraus N, McGee T. Electrophysiology of the human auditory system. In: Popper AN, Fay RR, editors. The Mammalian Auditory Pathway: Neurophysiology, vol. 2 New York: Springer-Verlag; 1992 (p. 335–403).

[38] Kraus N, McGee TJ, Carrell TD, Zecker SG, Nicol TG, Koch DB. Auditory neurophysiologic responses and discrimination deficits in children with learning problems. Science 1996;273(5277):971–3.

[39] Kraus N, McGee TJ, Koch DB. Speech sound perception and learning: biologic bases. Scand Audiol Suppl 1998:497–517.

[40] Meric C, Collet L. Visual attention and evoked otoacoustic emissions: a slight but real effect. Int J Psychophysiol 1992;12(3):233–5.

[41] Meric C, Collet L. Differential effects of visual attention on spontaneous and evoked otoacoustic emissions. Int J Psychophysiol 1994;17(3):281–9.

[42] Merzenich M, Wright B, Jenkins W, Xerri C, Byl N, Miller S, et al. Cortical plasticity underlying perceptual, motor, and cognitive skill development: implications for neurorehabilitation. Cold Spring Harb Symp Quant Biol 1996:611–8.

[43] Merzenich MM, Grajski K, Jenkins W, Recanzone G, Perterson B. Functional cortical plasticity. Cortical network origins of representational changes. Cold Spring Harb Symp Quant Biol 1991:55873–87.

[44] Merzenich MM, Jenkins WM, Johnston P, Schreiner C, Miller SL, Tallal P. Temporal processing deficits of language-learning impaired children ameliorated by training. Science 1996;271(5245):77–81.

[45] Miller GL, Knudsen EI. Adaptive plasticity in the auditory thalamus of juvenile barn owls. J Neurosci 2003;23(3):1059–65.

[46] Moore DR. Plasticity of binaural hearing and some possible mechanisms following late-onset deprivation. J Am Acad Audiol 1993;4(5):277–84.

[47] Moore DR. Auditory brainstem of the ferret: long survival following cochlear removal progressively changes projections from the cochlear nucleus to the inferior colliculus. J Comp Neurol 1994;339(2):301–10.

[48] Quian Quiroga R, Garcia H. Single-trial event-related potentials with wavelet denoising. Clin Neurophysiol 2003;114(2):376–90.

[49] Reed MA. Speech perception and the discrimination of brief auditory cues in reading disabled children. J Exp Child Psychol 1989;48(2):270–92.

[50] Rouiller E, Ribaupierre YD, Ribaupierre FD. Phase-locked responses to low-frequency tones in the medial geniculate-body. Hear Res 1979;1(3):213–26.

[51] Russo N, Musacchia G, Nicol T, ZeckerS, Kraus N. Brainstem responses to speech syllables. Clin Neurophysiol 2004.

[52] Schulte-Koerne G, Deimel W, Bartling J, Remschmidt H. The role of phonological awareness, speech perception, and auditory temporal processing for dyslexia. Eur Child Adolesc Psychiatry 1999;8:III/28–34.

[53] Starr A, Don M. Brain potentials evoked by acoustic stimuli. In: Picton T, editor. Human event-related potentials. Handbook of electroencephalography and clinical neurophsyiology, vol. 3 New York: Elsevier; 1988 (p. 97–158).

[54] Steinschneider M, Volkov IO, Noh MD, Garell PC, Howard III MA. Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. J Neurophysiol 1999;82(5):2346–57.

[55] Suga N, Gao E, Zhang Y, Ma X, Olsen JF. The corticofugal system for hearing: recent progress. Proc Natl Acad Sci USA 2000;97(22):11807–14.

[56] Suga N, Xiao Z, Ma X, Ji W. Plasticity and corticofugal modulation for hearing in adult animals. Neuron 2002;36(1):9–18.

[57] Tallal P, Miller SL, Bedi G, Byma G, Wang X, Nagarajan SS, et al. Language comprehension in language-learning impaired children improved with acoustically modified speech. Science 1996;271(5245):81–4.

[58] Taylor MM, Creelman CD. PEST: efficient estimates on probability functions. J Acoust Soc Am 1967;41:782–7.

[59] Tremblay K, Kraus N, McGee T. The time course of auditory perceptual learning: neurophysiological changes during speech–sound training. NeuroReport 1998;9(16):3557–60.

[60] Wallace MN, Rutkowski RG, Shackleton TM, Palmer AR. Phaselocked responses to pure tones in guinea pig auditory cortex. NeuroReport 2000;11(18):3989–93.

[61] Warrier CM, Johnson KL, Hayes EA, Nicol T, Kraus N. Learning impaired children exhibit timing deficits and training-related improvements in auditory cortical responses to speech in noise. Exp Brain Res 2004.

[62] Weinberger NM. Dynamic regulation of receptive fields and maps in the adult sensory cortex. Annu Rev Neurosci 1995;18:129–58.

[63] Wible B, Nicol T, Kraus N. Abnormal neural encoding of repeated speech stimuli in noise in children with learning problems. Clin Neurophysiol 2002;113(4):485–94.

[64] Wible B, Nicol T, Kraus N. Atypical brainstem representation of onset and formant structure of speech sounds in children with language-based learning problems. Biol Psychiatry 2004.

[65] Wilkinson G. Wide range achievement test-3. Wilmington, DE: Jastak; 1993.

[66] Woodcock R, Johnson M. Woodcock–Johnson Psycho-Educational Battery—revised. Tests of cognitive ability. Allen, TX: DLM Teaching Resources; 1977.

[67] Woodcock R, Johnson M. Woodcock–Johnson Psycho-Educational Battery—revised. Tests of cognitive ability. Allen, TX: DLM Teaching Resources; 1989.

[68] Wright BA, Lombardino LJ, King WM, Puranik CS, Leonard CM, Merzenich MM. Deficits in auditory temporal and spectral resolution in language-impaired children. Nature 1997;387(6629):176–8.

[69] Yan W, Suga N. Corticofugal modulation of time-domain processing of biosonar information in bats. Science 1996;273:1100–3.

[70] Zhang Y, Suga N. Modulation of responses and frequency tuning of thalamic and collicular neurons by cortical activation in mustached bats. J Neurophysiol 2000;84(1):325–33.