

Audio–vocal system regulation in children with autism spectrum disorders

Nicole Russo · Charles Larson · Nina Kraus

Received: 4 September 2007 / Accepted: 4 March 2008 / Published online: 18 March 2008
© Springer-Verlag 2008

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

N. Russo (✉) · C. Larson · N. Kraus
The Roxelyn and Richard Pepper Department of Communication Sciences, Northwestern University, Frances Searle Bldg, 2240 Campus Drive, Evanston, IL 60208, USA
e-mail: n-russo@northwestern.edu

N. Russo · C. Larson · N. Kraus
Northwestern University Interdepartmental Neuroscience Program, Northwestern University, Frances Searle Bldg, 2240 Campus Drive, Evanston, IL 60208, USA

N. Kraus
Department of Neurobiology and Physiology, Northwestern University, Frances Searle Bldg, 2240 Campus Drive, Evanston, IL 60208, USA

C. Larson · N. Kraus
Department of Otolaryngology, Northwestern University, Frances Searle Bldg, 2240 Campus Drive, Evanston, IL 60208, USA

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 and Peppe 2003; Rapin and 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 (Shriberg et al. 2001; Boucher 2003; Rapin and Dunn 2003; Siegal and 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 (Paul et al. 2005b; McCann et al. 2007).

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 (Khalifa et al. 2001). In contrast, Gravel et al. (2006) showed no behavioral differences in the peripheral auditory system in high-functioning children with autism. Tharpe et al. (2006) evaluated both peripheral audiometry and brainstem function in children with autism. 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 et al. (2006) did not find brainstem deficits, other studies of brainstem integrity have identified aberrant function (McClelland et al. 1992; Klin 1993; Maziade et al. 2000; Rapin and Dunn 2003; Rosenhall et al. 2003; Russo et al. 2008 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. 2008 in press). Further, there is ample evidence for deficient or atypical cortical processing of speech or speech-like stimuli associated with ASD (Wang et al. 2001; Boddaert et al. 2003, 2004; Ceponiene et al. 2003; Jansson-Verkasalo et al. 2003; Gervais et al. 2004; Kasai et al. 2005; Lepisto et al. 2005, 2006), including reports of deficient cortical processing specific

to prosody (Erwin et al. 1991; Wang et al. 2001; Kujala et al. 2005; Korpilahti et al. 2006). 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 (Leder et al. 1987; Perkell et al. 1992; Svirsky et al. 1992; Lane et al. 1997; Monini et al. 1997; Higgins et al. 1999; Hamzavi et al. 2000; Campisi et al. 2005). 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 (Lane and Tranel 1971) and sidetone amplification studies (Lane et al. 1961; Lane and Tranel 1971). 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) (Shriberg et al. 2001; Boucher 2003; Rapin and Dunn 2003; Siegal and 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 et al. 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 et al. 1998; Hain et al. 2000; Guenther 2006; 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 et al. 1998; Guenther 2006; Tourville et al. 2007). Further, they report that experimentation with speech begins early in development, as is evidenced by infant babbling. Hain et al. (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 (Lane and Tranel 1971). The concept of a "Theory of Mind" (Premack and 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 and Peppe 2003; 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 ≤ 20 dB for pure tone

octave frequencies 250–8,000 Hz), and a full scale mental ability confidence interval score ≥ 80 [Wechsler Abbreviated Scale Of Intelligence (WASI) (Woerner and Overstreet 1999)] (Table 1). 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 1) 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 5 s, 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 and 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

Table 1 Behavioral test scores

		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	7.890
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

Group mean 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. Mean and standard deviations are reported for these groups as well. 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

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 (Burnett et al. 1998; Hain et al. 2000; Bauer and Larson 2003); 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-s 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 and 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 three standard deviations (SD) of the prestimulus 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. 1). 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

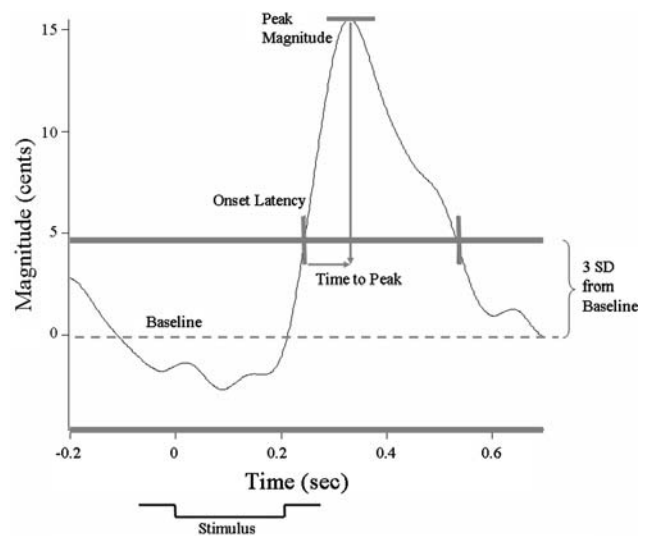


Fig. 1 Exemplar vocal response to pitch perturbation during auditory feedback. The stimulus (indicated underneath the x-axis) begins at time 0 and lasts 200 ms. To evaluate the pitch-shift reflex, first the baseline mean frequency (cents) and three standard deviations (SD) from the mean as threshold 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-meant to 0 cents

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 et al. 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) s; ASD = 0.23 (0.209) s]; time of peak magnitude [TD = 0.05 (0.015) s; ASD = 0.12 (0.101) s]; 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 2).

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 1.

Table 2 Pitch-shift reflex compensatory response measures

		Pitch-shift reflex measures		
		Onset latency (s)	Time to peak (s)	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

Group mean and standard deviations for compensatory response onset latency (s), time of peak magnitude (s), and magnitude (cents) are shown for TD and ASD groups. Mean and standard deviations are also reported for the ASD-LOW and ASD-HIGH subgroups

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. 2). 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 1) and indicated no differences in performance

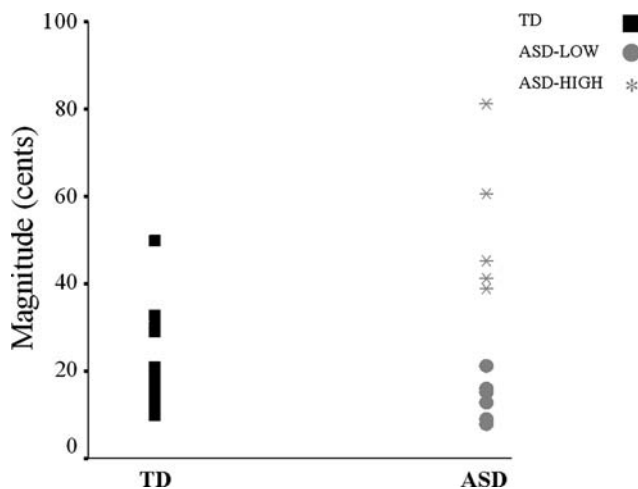


Fig. 2 Dot plot of compensatory response magnitudes (cents) of TD children (left; black squares) and children with ASD (right; ASD-LOW: gray circles and ASD-HIGH: asterisks). Children whose magnitudes were within ± 1.65 SD of the typical range comprise the ASD-LOW group, whereas children whose magnitudes exceeded $+1.65$ SD comprise the ASD-HIGH group

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. 3) 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) s, 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 2.

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 ($H(2) = 6.967$, $p = 0.031$).

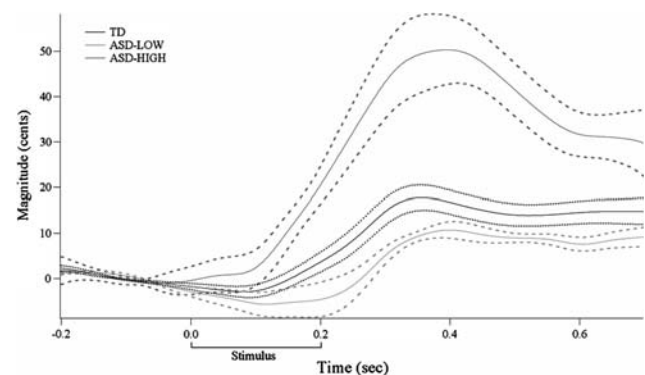


Fig. 3 Grand average TD (black), ASD-LOW (light gray) and ASD-HIGH (dark gray) response magnitude curves and standard error (TD: dotted lines; ASD: dashed lines). Whereas the TD response to pitch perturbation is approximately 20 cents, the ASD-LOW group response is significantly smaller, and the ASD-HIGH group response is significantly larger than the TD group

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 1.

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 ≤ 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 and Fidell 2007). 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. 4). 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. 5). 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

Fig. 4 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

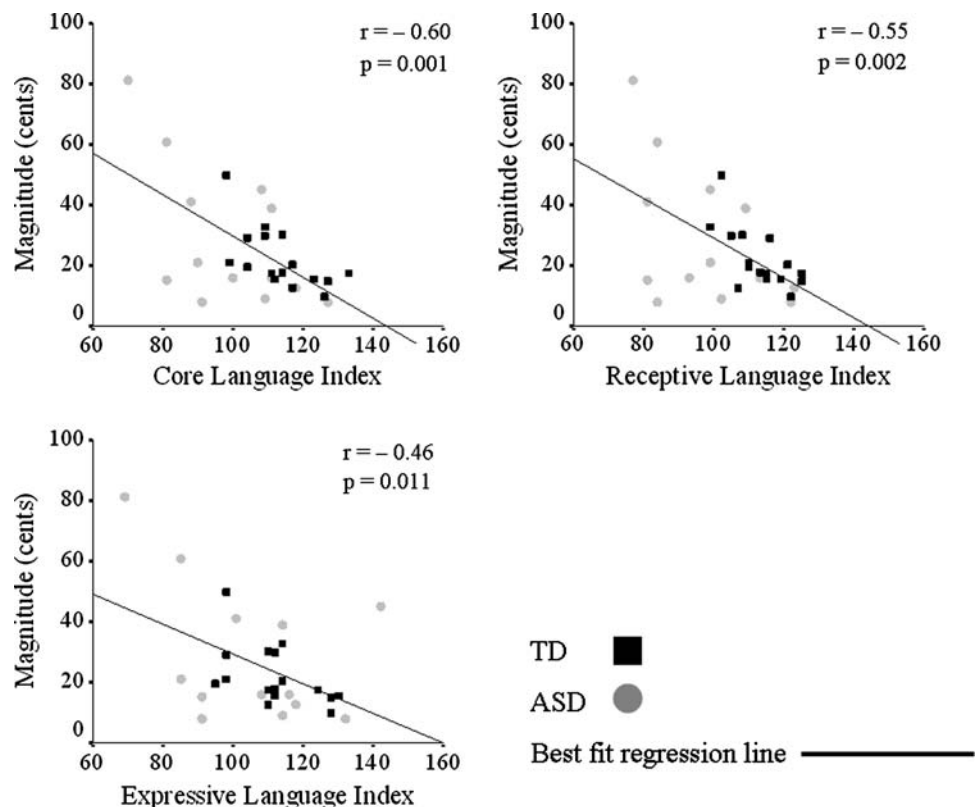
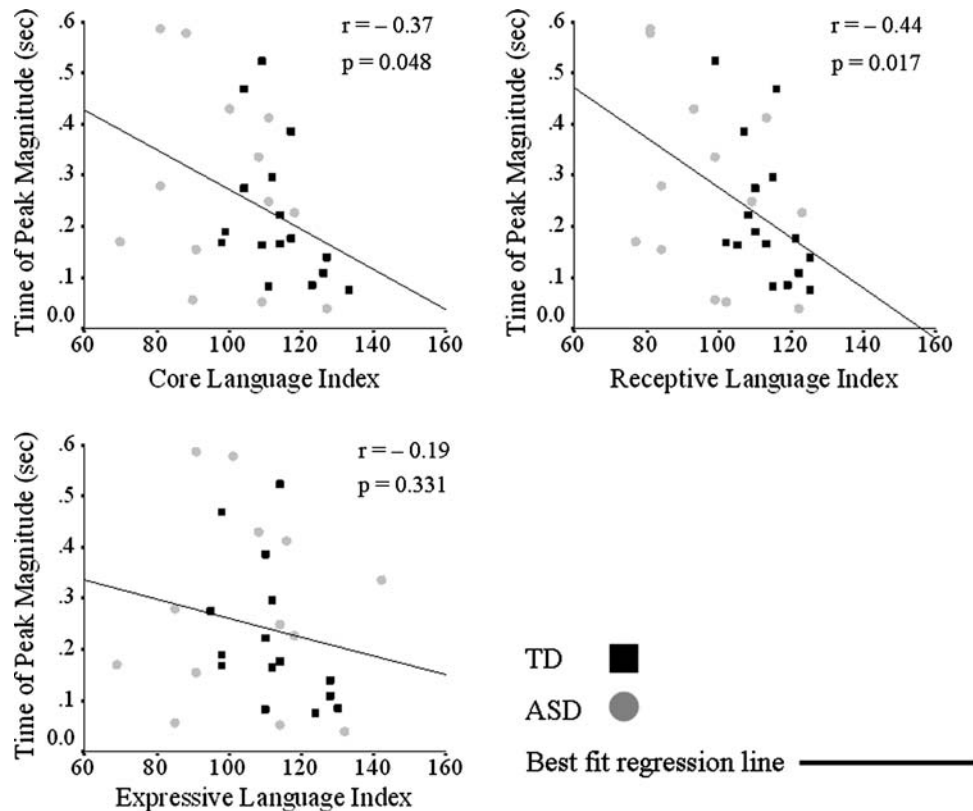


Fig. 5 Relationship between time of peak magnitude and language ability. Statistically significant Pearson's correlations ($r \geq 0.32$ and $p \leq 0.05$) existed between the time of peak magnitude (s) core (*top left*) and receptive (*top right*) language abilities, such that a shorter time to peak 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



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 eight 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 (Tharpe et al. 2006; Freitag 2007). 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 (Shriberg et al. 2001; McCann and Peppe 2003; Paul et al. 2005a, b). Moreover, ceiling effects are commonly noted in behavioral measures of prosody in ASD (Paul et al. 2005a, b). 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 and Larson 2007). 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 and 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 and Sherman 1983; Volman and 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 and Konishi 1999). In addition, 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 and 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). 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 et al. 2002) and non-human primates, such as the marmoset (Eliades and Wang 2003). Based on results from an auditory feedback/magnetoencephalography study, Houde et al. (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 (Siegal and Blades 2003; Herbert and Kenet 2007)], 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 and Pierce 2005). Given the known deficits in cortical processing of prosody in children with ASD (Erwin et al. 1991; Wang et al. 2001; Kujala et al. 2005; Korpilahti et al. 2006), 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 and 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 (O'Neill and Jones 1997; Khalifa et al. 2004; Kellerman et al. 2005). 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, 2006; Korpilahti 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 et al. 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 et al. 2001) and subcortical levels (Krishnan et al. 2004, 2005; Xu et al. 2006; Musacchia et al. 2007; Wong et al. 2007). 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 et al. 2006) and are less affected by alterations in auditory feedback (Zatorre et al. 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 15 min); it lends itself for use as an objective measure of one aspect of prosody deficits in ASD.

Acknowledgments 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.

References

- Baron-Cohen S, Knickmeyer RC, Belmonte MK (2005) Sex differences in the brain: implications for explaining autism. *Science* 310:819–823
- Bauer JJ, Larson CR (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* 114:1048–1054
- Boddaert N, Belin P, Chabane N, Poline JB, Barthelemy C, Mouren-Simeoni MC, Brunelle F, Samson Y, Zilbovicius M (2003) Perception of complex sounds: abnormal pattern of cortical activation in autism. *Am J Psychiatry* 160:2057–2060
- Boddaert N, Chabane N, Belin P, Bourgeois M, Royer V, Barthelemy C, Mouren-Simeoni MC, Philippe A, Brunelle F, Samson Y, Zilbovicius M (2004) Perception of complex sounds in autism: abnormal auditory cortical processing in children. *Am J Psychiatry* 161:2117–2120
- Boersma P, Weenink D (2004) PRAAT: doing phonetics by computer
- Boucher J (2003) Language development in autism. *Int J Pediatr Otorhinolaryngol* 67(Suppl 1):S159–163
- Burnett TA, Freedland MB, Larson CR, Hain TC (1998) Voice F_0 responses to manipulations in pitch feedback. *J Acoust Soc Am* 103:3153–3161
- Callan DE, Kent RD, Guenther FH, Vorperian HK (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* 43:721–736
- 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* 115:1046–1050
- Ceponiene R, Lepisto T, Shestakova A, Vanhala R, Alku P, Naatanen R, Yaguchi K (2003) Speech-sound-selective auditory impairment in children with autism: they can perceive but do not attend. *Proc Natl Acad Sci USA* 100:5567–5572
- Chen SH, Liu H, Xu Y, Larson CR (2007) Voice F_0 responses to pitch-shifted voice feedback during English speech. *J Acoust Soc Am* 121:1157–1163
- 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* 15:225–230
- Dawson G, Osterling J, Meltzoff AN, Kuhl P (2000) Case study of the development of an infant with autism from birth to two years of age. *J Appl Dev Psychol* 21:299–313
- Doupe AJ, Kuhl PK (1999) Birdsong and human speech: common themes and mechanisms. *Annu Rev Neurosci* 22:567–631
- Ejiri K (1998) Relationship between rhythmic behavior and canonical babbling in infant vocal development. *Phonetica* 55:226–237
- Eliades SJ, Wang X (2003) Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations. *J Neurophysiol* 89:2194–2207
- Erwin R, Van Lancker D, Guthrie D, Schwafel J, Tanguay P, Buchwald JS (1991) P3 responses to prosodic stimuli in adult autistic subjects. *Electroencephalogr Clin Neurophysiol* 80:561–571

- Filipek PA, Accardo PJ, Ashwal S, Baranek GT, Cook EH Jr, Dawson G, Gordon B, Gravel JS, Johnson CP, Kallen RJ, Levy SE, Minschew NJ, Ozonoff S, Prizant BM, Rapin I, Rogers SJ, Stone WL, Teplin SW, Tuchman RF, Volkmar FR (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* 55:468–479
- Freitag CM (2007) The genetics of autistic disorders and its clinical relevance: a review of the literature. *Mol Psychiatry* 12:2–22
- Gernsbacher MA (2004) Language is more than speech: a case study. *J Dev Learn Disord* 8:79–96
- Gervais H, Belin P, Boddaert N, Leboyer M, Coez A, Sfaello I, Barthelemy C, Brunelle F, Samson Y, Zilbovicius M (2004) Abnormal cortical voice processing in autism. *Nat Neurosci* 7:801–802
- Gravel JS, Dunn M, Lee WW, Ellis MA (2006) Peripheral audition of children on the autistic spectrum. *Ear Hear* 27:299–312
- Guenther FH (2006) Cortical interactions underlying the production of speech sounds. *J Commun Disord* 39:350–365
- Guenther FH, Hampson M, Johnson D (1998) A theoretical investigation of reference frames for the planning of speech movements. *Psychol Rev* 105:611–633
- Hain TC, Burnett TA, Kiran S, Larson CR, Singh S, Kenney MK (2000) Instructing subjects to make a voluntary response reveals the presence of two components to the audio–vocal reflex. *Exp Brain Res* 130:133–141
- Hamzavi J, Deutsch W, Baumgartner WD, Bigenzahn W, Gstoettner W (2000) Short-term effect of auditory feedback on fundamental frequency after cochlear implantation. *Audiology* 39:102–105
- Herbert MR, Kenet T (2007) Brain abnormalities in language disorders and in autism. *Pediatr Clin North Am* 54:563–583
- Higgins MB, McCleary EA, Schulte L (1999) Altered phonatory physiology with short-term deactivation of children's cochlear implants. *Ear Hear* 20:426–438
- Houde JF, Nagarajan SS, Sekihara K, Merzenich MM (2002) Modulation of the auditory cortex during speech: an MEG study. *J Cogn Neurosci* 14:1125–1138
- Hubbard K, Trauner DA (2007) Intonation and emotion in autistic spectrum disorders. *J Psycholinguist Res* 36:159–173
- Iarocci G, McDonald J (2006) Sensory integration and the perceptual experience of persons with autism. *J Autism Dev Disord* 36:77–90
- Jancke L, Gaab N, Wustenberg T, Scheich H, Heinze HJ (2001) Short-term functional plasticity in the human auditory cortex: an fMRI study. *Cogn Brain Res* 12:479–485
- Jansson-Verkasalo E, Ceponiemi R, Kielinen M, Suominen K, Jantti V, Linna S, Moilanen I, Naatanen R (2003) Deficient auditory processing in children with Asperger syndrome, as indexed by event-related potentials. *Neurosci Lett* 338:197–200
- Kasai K, Hashimoto O, Kawakubo Y, Yumoto M, Kamio S, Itoh K, Koshida I, Iwanami A, Nakagome K, Fukuda M, Yamasue H, Yamada H, Abe O, Aoki S, Kato N (2005) Delayed automatic detection of change in speech sounds in adults with autism: a magnetoencephalographic study. *Clin Neurophysiol* 116:1655–1664
- Kellerman GA, Fin J, Gorman JM (2005) Auditory abnormalities in autism: toward functional distinctions among findings. *CMS Spectr* 10:748–756
- Khalifa S, Bruneau N, Roge B, Georgieff N, Veuillet E, Adrien JL, Barthelemy C, Collet L (2001) Peripheral auditory asymmetry in infantile autism. *Eur J Neurosci* 13:628–632
- Khalifa S, Bruneau N, Roge B, Georgieff N, Veuillet E, Adrien JL, Barthelemy C, Collet L (2004) Increased perception of loudness in autism. *Hear Res* 198:87–92
- Klin A (1993) Auditory brainstem responses in autism: brainstem dysfunction or peripheral hearing loss? *J Autism Dev Disord* 23:15–35
- Korpilahti P, Jansson-Verkasalo E, Mattila ML, Kuusikko S, Suominen K, Rytty S, Pauls DL, Moilanen I (2006) Processing of affective speech prosody is impaired in Asperger syndrome. *J Autism Dev Disord* 37:1539–1549
- Krishnan A, Xu Y, Gandour JT, Cariani PA (2004) Human frequency-following response: representation of pitch contours in Chinese tones. *Hear Res* 189:1–12
- Krishnan A, Xu Y, Gandour J, Cariani P (2005) Encoding of pitch in the human brainstem is sensitive to language experience. *Cogn Brain Res* 25:161–168
- 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* 383:260–265
- Lane H, Tranel B (1971) The Lombard sign and the role of hearing in speech. *J Speech Hear Res* 14:677–709
- Lane HL, Catania AC, Stevens SS (1961) Voice level: autophonic scale, perceived loudness, and effects of sidetone. *J Acoust Soc Am* 33:160–167
- Lane H, Wozniak J, Matthies M, Svirsky M, Perkell J, O'Connell M, Manzella J (1997) Changes in sound pressure and fundamental frequency contours following changes in hearing status. *J Acoust Soc Am* 101:2244–2252
- Larson CR, Sun J, Hain TC (2007) Effects of simultaneous perturbations of voice pitch and loudness feedback on voice F_0 and amplitude control. *J Acoust Soc Am* 121:2862–2872
- Leder SB, Spitzer JB, Milner P, Flevaris-Phillips C, Kirchner JC, Richardson F (1987) Voice intensity of prospective cochlear implant candidates and normal hearing adult males. *Laryngoscope* 97:224–227
- Leonardo A, Konishi M (1999) Decrystallization of adult birdsong by perturbation of auditory feedback. *Nature* 399:466–470
- Lepisto T, Kujala T, Vanhala R, Alku P, Huottilainen M, Naatanen R (2005) The discrimination of and orienting to speech and non-speech sounds in children with autism. *Brain Res* 1066:147–157
- 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* 117:2161–2171
- Liu H, Larson CR (2007) Effects of perturbation magnitude and voice F_0 level on the pitch-shift reflex. *JASA* 122:3671–3677
- 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* 24:659–685
- Lord C, Risi S, Lambrecht L, Cook EH Jr, Leventhal BL, DiLavore PC, Pickles A, Rutter M (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* 30:205–223
- 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* 18:199–211
- Margoliash D (2002) Evaluating theories of bird song learning: implications for future directions. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 188:851–866
- Marler P, Sherman V (1983) Song structure without auditory feedback: emendations of the auditory template hypothesis. *J Neurosci* 3:517–531
- Maziade M, Merette C, Cayer M, Roy MA, Szatmari P, Cote R, Thivierge J (2000) Prolongation of brainstem auditory-evoked

- responses in autistic probands and their unaffected relatives. *Arch Gen Psychiatry* 57:1077–1083
- McCann J, Peppe S (2003) Prosody in autism spectrum disorders: a critical review. *Int J Lang Commun Disord* 38:325–350
- McCann J, Peppe S, Gibbon FE, 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* 42:682–702
- McClelland RJ, Eyre DG, Watson D, Calvert GJ, Sherrard E (1992) Central conduction time in childhood autism. *Br J Psychiatry* 160:659–663
- Miller CA (2006) Developmental relationships between language and theory of mind. *Am J Speech Lang Pathol* 15:142–154
- Monini S, Banci G, Barbara M, Argiro MT, Filipo R (1997) Clarion cochlear implant: short-term effects on voice parameters. *Am J Otol* 18:719–725
- 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 USA* 104:15894–15898
- Natke U, Donath TM, Kalveram KT (2003) Control of voice fundamental frequency in speaking versus singing. *J Acoust Soc Am* 113:1587–1593
- O'Neill M, Jones RS (1997) Sensory-perceptual abnormalities in autism: a case for more research? *J Autism Dev Disord* 27:283–293
- Paul R, Augustyn A, Klin A, Volkmar FR (2005a) Perception and production of prosody by speakers with autism spectrum disorders. *J Autism Dev Disord* 35:205–220
- Paul R, Shriberg LD, McSweeney J, Cicchetti D, Klin A, Volkmar F (2005b) Brief report: relations between prosodic performance and communication and socialization ratings in high functioning speakers with autism spectrum disorders. *J Autism Dev Disord* 35:861–869
- Perkell J, Lane H, Svirsky M, Webster J (1992) Speech of cochlear implant patients: a longitudinal study of vowel production. *J Acoust Soc Am* 91:2961–2978
- Prather JF, Peters S, Nowicki S, Mooney R (2008) Precise auditory-vocal mirroring in neurons for learned vocal communication. *Nature* 451:305–310
- Premack D, Woodruff G (1978) Does the chimpanzee have a theory of mind? *Behav Brain Sci* 4:515–526
- Rapin I, Dunn M (2003) Update on the language disorders of individuals on the autistic spectrum. *Brain Dev* 25:166–172
- Rosenhall U, Nordin V, Brantberg K, Gillberg C (2003) Autism and auditory brain stem responses. *Ear Hear* 24:206–214
- Russo N, Skoe E, Trommer B, Nicol T, Zecker S, Bradlow A, Kraus N (2008) Deficient brainstem encoding of pitch in children with autism spectrum disorders. *Clin Neurophysiol* (in press)
- Sankoh AJ, Huque MF, Dubey SD (1997) Some comments on frequently used multiple endpoint adjustment methods in clinical trials. *Stat Med* 16:2529–2542
- Semel E, Wiig EH, Secord WA (2003) Clinical evaluation of language fundamentals, 4th edn. In: Harcourt assessment, Inc., San Antonio, TX
- Shriberg LD, Paul R, McSweeney JL, Klin AM, Cohen DJ, Volkmar FR (2001) Speech and prosody characteristics of adolescents and adults with high-functioning autism and Asperger syndrome. *J Speech Lang Hear Res* 44:1097–1115
- Siegal M, Blades M (2003) Language and auditory processing in autism. *Trends Cogn Sci* 7:378–380
- Svirsky MA, Lane H, Perkell JS, Wozniak J (1992) Effects of short-term auditory deprivation on speech production in adult cochlear implant users. *J Acoust Soc Am* 92:1284–1300
- Tabachnick B, Fidell L (2007) Principal components and factor analysis. In: Hartman S (ed) Using multivariate statistics. Pearson Education, Boston, pp 607–675
- Tharpe AM, Bess FH, Sladen DP, Schissel H, Couch S, Schery T (2006) Auditory characteristics of children with autism. *Ear Hear* 27:430–441
- Titze IR (1994) Fluctuations and perturbations in vocal output. In: Principles of voice production. Prentice Hall, Englewood Cliffs, pp 279–306
- Tourville JA, Reilly KJ, Guenther FH (2007) Neural mechanisms underlying auditory feedback control of speech. *NeuroImage* 39:1429–1443
- Volman SF, Khanna H (1995) Convergence of untutored song in group-reared zebra finches (*Taeniopygia guttata*). *J Comp Psychol* 109:211–221
- Wang A, Dapretto M, Hariri A, Sigman M, Bookheimer S (2001) Processing affective and linguistic prosody in autism: an fMRI study. *NeuroImage* 13:621
- Woerner C, Overstreet K (eds) (1999) Wechsler abbreviated scale of intelligence (WASI). The Psychological Corporation, San Antonio
- Wong P, Skoe E, Russo N, Dees T, Kraus N (2007) Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat Neurosci* 10:420
- Xu Y, Krishnan A, Gandour JT (2006) Specificity of experience-dependent pitch representation in the brainstem. *Neuroreport* 17:1601–1605
- Zatorre RJ, Chen JL, Penhune VB (2007) When the brain plays music: auditory-motor interactions in music perception and production. *Nat Rev Neurosci* 8:547–558