



Published in final edited form as:

Res Autism Spectr Disord. 2008 January 1; 2(1): 110–124. doi:10.1016/j.rasd.2007.04.001.

Production of Syllable Stress in Speakers with Autism Spectrum Disorders

Rhea Paul, Ph.D.,

Southern Connecticut State University, Yale Child Study Center

Nancy Bianchi, M.S.,

West Haven, CT Public Schools

Amy Augustyn, M.S.,

Florida State University

Ami Klin, Ph.D., and

Yale Child Study Center

Fred Volkmar, M.D.

Yale Child Study Center

Keywords

syllable stress; prosody; autism

Autism Spectrum Disorders (ASDs) are a group of severe neuropsychiatric conditions characterized by disturbances in social, cognitive, and communicative function that are not fully explained by developmental level. Although most of these disorders are associated with depressed cognitive and language functioning, an estimated 20%–40% of individuals with these syndromes function within the normal range on IQ testing (American Psychiatric Association, 1994; Klin & Volkmar, 1997). These individuals demonstrate large spoken vocabularies and apparently intact formal language skills. Individuals at this level of functioning may, according to the Diagnostic and Statistical manual of Mental Disorders-4th Ed.-TR (APA, 2000), receive one of three diagnoses within the autism spectrum: high functioning autism (ASD), in which there is a history of language delay and symptoms in all three areas that characterize the syndrome (severe deficits in socialization, communication and stereotyped, repetitive or ritualistic behaviors); Asperger syndrome (AS) in which there is no history of language delay, the presence of significant social and communicative disability and an obsessive interest in circumscribed topics; and PDD-NOS in which social, communicative and/or stereotypic behaviors are present, but do not reach criteria for autism. The most prominent communication deficits in these disorders in higher functioning individuals are in the areas of pragmatics and social communication (Ramberg, Ehlers, Nyden, Johansson, & Gillberg, 1996; Tager-Flusberg, 1995). Another area in which communicative difficulties are frequently reported for speakers with ASD is *prosody* (McCann & Peppe, 2003).

Address for correspondence: Rhea Paul, Ph.D., Yale Child Study Center, 230 S. Frontage Rd., New Haven, CT 06520-7900203, Phone: 785 3388; Fax: 203 737 4197, Email: rhea.paul@yale.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

The term *prosody* refers to the suprasegmental aspects of speech production; those properties of the speech signal that extend beyond phonemic segments to modulate and enhance its meaning (Crystal, 1969; Couper-Kuhlen, 1986; Kent & Read, 1992; Merewether & Alpert, 1990; Panagos & Prelock, 1997). The term *prosody* is typically used to refer to:

1. the assignment of relative prominence or *stress* to various units within the signal
2. changes in pitch of the speech sound wave over time that make up its *intonation contour*;
3. the rhythm and timing patterns that make up the *phrasing* of the utterance; expressed through *rate*, *duration* and *pauses* within speech events;

(Lehiste, 1970; Shriberg, Kwiatkowski, & Rasmussen, 1990). Acoustically, prosody is a composite of pitch (fundamental frequency), intensity (amplitude), and duration, as well as the co-variation of these variables (Stephens, Nickerson, & Rollins, 1983).

Since the first delineation of the autistic syndrome (Kanner, 1943), abnormal prosody has been frequently identified as a core feature of the syndrome for individuals with autism who speak (Baltaxe & Simmons, 1987, 1992; Fay & Schuler, 1980; Ornitz & Ritvo, 1976; Paul, 1987; Pronovost, Wakstein, & Wakstein, 1966; Rutter & Lockyer, 1967; Tager-Flusberg, 1981). Paul, Shriberg, et al. (in press) reported abnormal prosody in 47% of the speakers with ASD studied. These abnormalities have been reported anecdotally to include monotonic or machine-like intonation, deficits in the use of pitch and control of volume, deficiencies in vocal quality, and use of aberrant stress patterns.

Shriberg et al., (2001) reported on a range of suprasegmental characteristics of continuous speech in speakers with ASD, using a standard assessment method, the Prosody-Voice Screening Profile (PVSP; Shriberg, Kwiatkowski, & Rasmussen, 1990). This study found significant prosodic differences between speakers with ASD and typical speakers. However, differences were not wide-spread, but focused in a few areas; most notably in speech phrasing or fluency, the presence of hypernasal voice quality, and in the use of stress.

Stress, or the highlighting of particular words or syllables with increased duration, pitch changes, and amplitude (volume), is used for a variety of purposes in speech. In English, one function of stress is to distinguish grammatical class in some disyllabic words. For example in the word *present*, pronunciation with stress on the first syllable denotes a noun (**pre**' sent), while stress on the second denotes a verb (pre **sent**'). This function would generally be considered a grammatical usage of stress, since the prosodic change is employed to signal a change in grammatical class (Quirk, R., Greenbaum, S., Leech, G., & Svartvik, 1990). Another function that can be served by stress is the contrastive or emphatic function. This usage of stress involves highlighting a particular word within a sentence to mark it as salient or to point out its contrast with a previous element in the discourse. For example:

Speaker A: I need a red pencil.

Speaker B: I have a **blue** one.

This use of stress is generally considered a pragmatic function (Halliday, 1975), as it serves to focus attention on an aspect of the discourse that the speaker intends to mark as new or important. Chafe (1970) has argued that languages contain devices used not only to encode meaning but also to point out which constituents refer to material that should be foregrounded in consciousness. One of these devices for foregrounding is emphatic or contrastive stress (Solon, 1980).

Earlier studies of the production of stress in speakers with ASD have yielded mixed results. Baltaxe (1984) found more misassignments of contrastive stress in speakers with HFA than in typical controls, although the total number of misassignments was small. Baltaxe and Simmons (1987) reported that speakers with ASD misassigned stress to function rather than content words and used more than one stress within a sentence. Similarly, McCaleb and Prizant (1985) found that speakers with ASD did not use stress accurately to mark new and given information within sentences. Baltaxe and Guthrie (1987) showed that speakers with ASD made more errors than typical age-mates in grammatical placement of stress, although errors were made by both groups. However, Fay (1969) showed the speakers with HFA recalled stressed words better than unstressed ones, as typical speakers do. Fine, Bartolucci, Ginsberg, and Szatmari (1991) reported that speakers with ASD were similar to speakers with TD in their use of grammatical stress. Foreman (2002) looked at production and perception of emphatic stress in experimental paradigms, using acoustic analysis of production. She reported that high functioning individuals with ASD performed more accurately on perception than production tasks, whereas subjects with TD showed the opposite pattern. Overall, subjects with TD had better performance on production tasks, but both groups were similar on perception. Paul, Augustyn, Klin, and Volkmar (2005) however, found that adolescents with ASD were significantly less able than peers with TD to perform accurately on experimental tasks of both perception and production of emphatic stress. Thus, the current literature on the role of stress in the communicative competence of speakers with ASD remains contradictory.

The present study investigates the ability of adolescent speakers with ASD and typically developing (TD) age mates to imitate stress in nonsense syllables, using a standard measure, the Tennessee Test of Rhythm and Intonation (TTRIP; Koike & Asp, 1981). The rationale for employing a nonsense-syllable repetition task is to establish whether or not, at the simplest level in a nonmeaningful context, speakers with ASD are different from TD speakers in the ability to produce the perceptual and acoustic parameters of stressed syllables, with confounding variables, such as grammatical or pragmatic function and executive planning demands, removed. Our hypothesis is that in this reduced-demand context, speakers with ASD will show no differences from TD speakers. If the hypothesized result is borne out by the data, it would be reasonable to conclude that any difficulties in the meaningful, contextualized use of stress observed in speakers with ASD could not be attributed to any low-level deficits in the perception of stress differences in speech stimuli or to the neuromotor abilities involved in producing syllables perceived as stressed. If, however, there are differences observed on this simplified speech task, more fundamental processes of speech perception and/or neuromotor coordination may underlie some of the perceived oddities of prosody in speakers with ASD.

Method

Participants

Speech samples were collected for 66 subjects; 46 with autism spectrum disorders (ASD) and 20 with typical development (TD). These participants with ASD comprised all the individuals with any form of ASD who were seen in conjunction with ongoing studies of high functioning autism at the Yale Child Study Center's Developmental Disabilities Section within a two year time period. To qualify for the high functioning study, subjects were required to have a diagnosis within the ASD spectrum, a verbal IQ of 70 or greater and fluent use of spoken language. These individuals had all completed an extensive diagnostic protocol as part of two projects on the neurobiology of autism. The protocol included data from standardized assessments of cognitive (Wechsler scales), language (Clinical Evaluation of Language Fundamentals [CELF]-III, Semel, Wiig, & Secord, 1995), and social-adaptive functioning (Vineland Adaptive Behavior Scales, Sparrow, Balla, & Cichetti, 1984), and a videocassette

recording of a conversational speech sample obtained during a semistructured diagnostic interview.

The subjects with ASD, 43 males and 3 females, ranged in age from 7 years, 4 months to 28 years, 7 months. Diagnostic characterization included the Autism Diagnostic Interview-Revised (ADI-R; Lord, Rutter, & LeCouteur, 1994) and the Autism Diagnostic Observation Schedule-Generic (ADOS-G; Lord et al., 2000). Diagnostic assignment followed DSM-IV-TR criteria for HFA, AS, and Pervasive Developmental Disorders-Not Otherwise Specified (American Psychiatric Association, 2000). Clinical diagnoses were confirmed independently by two experienced clinicians (AK and FV) with demonstrated interrater reliability (Klin, Lang, Cicchetti, & Volkmar, 2000). Forty-eight percent (22) of the subjects were diagnosed as HFA by these methods; 41% (19) as AS, and 11% (5) as Pervasive Developmental Disorders-Not Otherwise Specified (PDD-NOS). As can be seen in Table 1, subjects' average age was 13.2 years (s.d. 4.4).

There were no differences among the three ASD diagnostic groups in terms of their average age, Performance IQs, functional communication scores or functional social adaptation level on the Vineland Adaptive Behavior Scales (Sparrow, Balla, & Cicchetti, 1984), suggesting all participants had similar difficulties in adaptive communication and social skills, despite their average IQ scores. There were significant differences, which reflect diagnostic assignment, on other measures that characterize the participants with ASD. On Verbal IQ and CELF Expressive Communication scores, the group with AS scored significantly higher than those with HFA or PDD-NOS. Individuals with AS in this sample, as well as those in the general population, typically have higher verbal IQs and a fluent and verbose speech style that is reflected in this difference, and forms part of the diagnostic criteria for AS (Klin, Volkmar, & Sparrow, 2000). The significant differences seen on the ADOS Communication and Socialization scores similarly reflect diagnostic criteria. Subjects who receive a diagnosis of HFA typically show more severe symptoms of the syndrome than those with either AS or HFA; individuals typically receive a diagnosis of PDD-NOS if they show relatively low levels of autistic symptomatology, which is reflected in the distribution of ADOS scores seen in this sample.

TD subjects included 17 males and 3 females who ranged in age from 7 years, 11 months to 27 years, 5 months. The comparison group was recruited from local schools and colleges through personal invitations. All were enrolled in appropriate grade for age in school or University, were considered to be achieving normally, and had no history of speech, language, or learning problems or of special education.

Procedures

Subjects were seen individually by the first and third authors (RP, AA) to complete the Tennessee Test of Rhythm and Intonation Patterns (T-TRIP), after procedures for the other ongoing research protocols had been completed. The T-TRIP consists of 25 items prerecorded on audiotape that vary in rhythm and intonation using the same nonsense syllable /ma/, as shown in Figure 1.

This study examined responses only to the first section of the test, in which test items 1–14 are systematically varied in stress pattern and number of syllables (two to six). Each of the 14 items was played from an audiotape obtained from the authors (Koike & Asp, 1981) on a Califone Model 1300AV audio cassette recorder set at a standard volume. The tape was stopped by the examiner after each item and the subject, after a series of five practice items, was asked to imitate the string of syllables heard. The subject spoke responses into a Shure Model SM10A dynamic head-mounted microphone, which recorded responses into a Marantz Model PMD222 portable cassette recorder, using one Maxell audiotape per subject.

Perceptual Ratings—A research assistant blind to the subjects' diagnostic assignment, and to the model presented to the subject, then rated each syllable string produced, and assigned each syllable within the string to either the stressed or unstressed category. A second rater, similarly blind, also rated a randomly selected 15% sample of the audiotapes, 7 from the ASD group and 3 from the TD group. Point-to-point reliability between the two sets of perceptual ratings was 96%.

Instrumental Measures—Each subject's recorded data were then analyzed by the second author (NB), who played the audio signal into the Kay Elemetrics Computerized Speech Lab program, Model 4300 (Kay Elemetrics, Lincoln Park, N.J.), which was used to measure the pitch range and duration of each stressed and unstressed syllable on the recording. The duration and frequency contour of the syllables were measured instrumentally, as described in Snow (2001a) and Schwartz, Petinou, Goffman, Lazowski, and Cartusciello (1996).

Figure 2 represents the time waveform and frequency contour of the utterance “**ma**-ma” generated by the CSL program, with the cursor marking off the stressed syllable. The units of analysis were each instance of the production in /ma/, in stressed or unstressed form. For each subject there were 32 data points for unstressed and 32 data points for stressed syllables; thus, there was a total of 64 data points. For each data point, six variables were measured: (1) duration in stressed syllables, (2) duration in unstressed syllables, (3) high pitch in stressed syllables, (4) low pitch in stressed syllables, (5) high pitch in unstressed syllables, and (6) low pitch in unstressed syllables. The boundaries of each syllable were set at the amplitude peak of the first or last periodic cycle that was visually distinct in the time waveform (the display of the signal amplitude over time; See Figure 2).

Duration measure—The duration of the nonsense syllable /ma/ was defined as the difference between the beginning and ending time boundaries of the syllable. The onset of the syllable was identified by the presence of glottal pulses of the second formant (F2) in the wideband display and a concomitant increase in amplitude on the combination display. The syllable end points were identified as the cessation of glottal pulses in F2 and a simultaneous drop in the amplitude display. The investigator manually controlled the cursors to mark off the boundaries of each syllable to measure duration and pitch. Duration in milliseconds was then automatically calculated by the CSL program for each of the 64 (32 stressed, 32 unstressed) syllables presented in the first 14 T-TRIP items for each subject.

Fundamental frequency and accent range measure—The automatic pitch extraction algorithm of the CSL program generated the fundamental frequency contour between the voicing boundaries. The investigator captured the fundamental frequency contour by manually controlling the cursors to measure the highest and lowest pitch of each syllable. The accent range is a ratio of the highest versus the lowest pitch in each stressed and unstressed syllable converted to octave units (one octave = 12 semitones = 1200 cents). This conversion to the octave scale adjusted the fundamental frequency data to approximate perceptually equivalent units (Burns & Ward, 1982). The formula to compute the accent or octave range (Y) starts with the relationship between the high fundamental frequency (H) and the low fundamental frequency (L) within each syllable. The following equation, as solved, yields the octave range in semitones:

$$Y=12/\log (2) \times \log (H/L) .$$

Accent range was recorded for each of the 64 syllables presented in the first 14 T-TRIP items for each subject.

Reliability of Instrumental Measures—To determine inter-rater agreement, a research assistant was trained in the CSL program, and acted as the second rater. The second rater re-analyzed 7 (10% of the total sample) randomly selected T-TRIP recordings for interscorer reliability. Reliability was calculated by determining the variation between the first rater's measure of the 7 subjects' sample set, and the second rater's measure. Reliability was expressed as a percentage agreement. This variation was calculated by dividing the two raters' values and multiplying by 100. The average percent agreement for each measure for each syllable from each subject was calculated and summarized. From these averages, the overall mean inter-rater percent agreement was found to be 96%.

RESULTS

Correlational analysis was used to determine whether there were relationships among any of the subjects' diagnostic characteristics (e.g., PIQ, VIQ, Vineland and ADOS scores) and results of the analysis of stress production. The only variable that showed any significant correlation with measures of stress production was Verbal IQ ($r = -.33$ and $-.45$ with stressed and unstressed accent range, respectively). For this reason, all analyses were done using VIQ as a covariate, to control for its effect on results.

Perceptual ratings

Analysis of variance was used, first, to determine whether there were differences among the three ASD groups in terms of their perceptual ratings on stress production. A difference was found for stressed syllables ($F = 3.3$; $p < 0.4$); therefore, analysis was run among the four groups (HFA, AS, PDD, TD). The Repeated Measures analysis of co-variance (ANCOVA) procedure of the General Linear Model program of the SPSS 11.5 Program was used to conduct this analysis. VIQ was entered as a covariate in the analysis, diagnostic group was the between subjects variable with four levels (HFA, AS, PDD-NOS, TD), and stress the within subjects variable with two levels (stressed and unstressed syllable assignments). Table 2 displays the average percentage of syllables produced appropriately as stressed or unstressed by each diagnostic group, as judged by listeners blind to the subjects' diagnostic category. ANCOVA revealed no significant within subject effects, suggesting there were not differences in the percent accuracy of stressed vs. unstressed syllables. There was, however, a significant between-subjects effect ($F = 7.3$; $p < .001$) when controlling for VIQ. Planned contrasts, using the Least Significant Difference procedure, were examined between the TD group and the other three diagnostic groups (See Table 2). These revealed that for both stressed and unstressed syllables, subjects were HFA were significantly less likely to receive correct ratings on their productions. There was no difference between the AS and TD groups. The PDD-NOS group was different from the TD in the stressed condition only. It can also be noted that there was a larger degree of variability in the two groups that were significantly different from the TDs. Standard deviations for the HFA and PDD-NOS group showed a wider range of variation than for the other two groups, with accuracy rates for some subjects as low as 66% stressed syllables, and as low as 52% for unstressed syllables.

Instrumental Measures

Two repeated measures ANCOVAs were used to determine differences in acoustic measures of stress production among the four diagnostic groups. Data for these analyses appear in Table 3. These data were submitted to two separate repeated measures ANCOVAs with one between-subjects variable with four levels (diagnosis: HFA, AS, PDD-NOS, TD) and one within-subjects variable in each analysis; Duration in stressed vs. unstressed syllables in one analysis, and Accent Range in stressed vs. unstressed syllables in the other. Again VIQ was used as a covariate in both analyses.

For the Duration analysis, there was no overall effect of diagnostic group ($F=.14$; $p < .26$); however, there was a significant main effect for stress, with overall durations in stressed syllables longer than those in unstressed ($F=5.6$; $p < .02$) for all diagnoses. There was also a significant duration \times diagnosis interaction ($F=4.8$; $p < .005$), suggesting that there was less difference between stressed and unstressed syllable durations in some groups than in others. To explore this interaction, and because there were no between-diagnostic group differences in this analysis, the three ASD groups were combined, with values for stressed and unstressed duration averaged across all three groups. These were then contrasted with durations for the TD group. This comparison appears in Figure 3. As can be seen there, both groups produced longer durations for stressed than unstressed syllables, but the difference between stressed and unstressed durations was greater for subjects with TD than for those with ASD. ANCOVA testing of the combined ASD group contrasted to the TD replicated this result. It found the significant main effect for duration ($F=711$; $p < .001$), but not for diagnostic group ($F=.76$; $p < .39$), and a significant duration \times diagnosis interaction effect ($F=10.3$; $p < .02$), as well. Subjects with TD had average durations of 346 msec. (s.d.: 44 msec.) in stressed syllables, 186 msec. (s.d.: 23 msec.) in unstressed syllables; those with ASD had durations of 321 msec. (s.d.: 45 msec.) in stressed syllables, 196 msec. (s.d.: 35 msec.) in unstressed. Figure 3 presents the data on duration in stressed and unstressed syllables in these two diagnostic groups.

The Accent Range (AR) ANCOVA yielded no comparisons that reached significance, when controlling for VIQ, either contrasting the four diagnostic groups, or comparing only the combined ASD group to TD. Averaged values across the ASD groups contrasted to those for the TD group are displayed in Figure 4. TD speakers had average accent ranges of 5.69 semitones in stressed syllables (s.d.: 2.07) and 4.52 semitones in unstressed syllables (s.d.: 1.77); those with ASD had higher accent ranges than those with TD for both kinds of syllables: 6.32 semitones in stressed syllables (s.d.: 2.46) and 5.10 in unstressed syllables (s.d.: 2.10). This trend failed to reach significance, however.

DISCUSSION

The findings of this analysis of production of stress in nonsense syllables by speakers with autism revealed that there were small but significant differences in the ability to produce syllables perceived by listeners as stressed and unstressed in this imitation task. We also found greater variability in the perceptual analysis for subjects with HFA and PDD-NOS than those with TD or AS. Instrumental analysis revealed that there were no significant differences in the pitch ranges produced within nonsense syllables among diagnostic groups; all speakers produced greater accent ranges in stressed than unstressed syllables. Accent ranges were consistently larger in both stressed and unstressed syllables for speakers with ASD, but this difference failed to reach significance. There were significant differences in the duration of stressed and unstressed syllables for all groups, as well, and there was also a significant interaction effect in the ANCOVA examining duration, such that the ASD group showed significantly less difference between stressed and unstressed syllable durations than the TD speakers did.

Our hypothesis, that speakers with ASD would show no differences from typical speakers on this nonsense imitation task, failed to be fully supported in this study. Unfortunately, then, its results do not allow us to fully resolve the issue regarding competencies with stress production in speakers with ASD. The small but significant differences seen here in both listener judgments and duration analyses may indeed suggest an underlying difficulty in the perceptual and/or motor apparatus involved in stress production.

However, it is also possible that the differences observed could be explained by social factors. For example, speakers with ASD may have depressed motivation to attend to and maximize

performance in order to succeed at this rather uninteresting task. Anecdotally, this was not our impression; in fact we observed that speakers with ASD seemed to be concentrating and trying very hard, whereas the typical young people who participated thought the task somewhat silly and did not appear to be working very diligently at it, seeming instead to perform it to some degree on “automatic pilot.”

What does seem clear from these data, however, is that difficulties in comprehending and managing the conversational functions of stress production cannot fully account for differences in use of prosodic stress that have been reported in the literature for speakers with ASD (e.g., Baltaxe, 1984; Baltaxe & Simmons, 1987; Baltaxe & Guthrie, 1987; McCaleb & Prizant, 1985; Shriberg et al., 2001). Even in a very simple, nonmeaningful task in which no linguistic or pragmatic value is associated with the stressed/unstressed distinction, this study suggests speakers with ASD, particularly those with HFA and PDD-NOS, show subtle difficulties on both perceptual and acoustic measures. Whether these difficulties are better accounted for by motor, perceptual, or social deficits is at this point a matter of speculation. There is, however, some support in the literature for hypothesizing that social factors may play a role. For instance, both Shriberg et al. (2001) and Gibbon, McCann, Peppe, O'Hare, & Rutherford (2004, September) reported an unusually high prevalence of residual speech errors in speakers with ASD, with rates at 20–30%, as opposed to the 1% seen in the general population (National Health Statistics on Voice, Speech, and Language, National Institute of Deafness and Communication Disorders website: <http://www.nidcd.nih.gov/health/statistics/vsl.asp>). Again, the reason for this excess cannot be fully determined, but one hypothesis involves the notion of social emulation. In order to develop from a distorted /s/, /r/, or /l/ to a more precise articulation, a young speaker needs to attend closely to models in the environment (what Shriberg [1987, October] called “tuning in”) and make small and careful adjustments in one’s own production to match these models (what Shriberg [1987, October] called “tuning up.”) The ability and motivation to tune in and tune up; that is, to focus on subtle details of visual and auditory aspects of the speech of others and to make the minute adjustments in one’s own production by attending to feedback from one’s own utterances in order to emulate others’ speech, may be less well-developed in speakers with ASD. This drive toward social emulation, which motivates typical children to tune in and tune up in order to sound like the speakers in their community may be one aspect of the social communicative deficit in speakers with ASD.

Another line of research that lends support to this suggestion has shown that children with ASD are less likely than their siblings with typical development to acquire the accent of their peers when they grow up in a non-English speaking household within an English-speaking country (Baron-Cohen & Stauton, 1995). Again, a failure to emulate significant speech peers in the environment is reflected in these findings. If children with ASD show less social emulation of details of speech production, a similar mechanism could be operating in the production of stress. Like speech distortions, the differences in stress differentiation in speakers with ASD in the current study do not “make or break” their utterances. Their stress is perceived appropriately much of the time, and acoustic differences are more a matter of quantity than quality. Nonetheless, their stress production is not quite at the level of their peers. This small difference, like their speech distortions, will not have huge consequences for their ability to communicate wants and needs. Rather, it will result merely in their speech sounding subtly “off” to others, which may, because of their reduced social motivation and awareness, not make a great deal of difference to the speaker with ASD. As a result, these subtle differences, which typical children would “tune up” over time, become habitual and persist.

If this speculation contains any kernel of truth, it could have some implications for treating the residual speech and prosody differences often seen in this population. It suggests that the type of intervention needed would focus not on the perceptual and motor details of the particular speech or prosodic difference, but more globally on the concept of comparing one’s speech

production to that of peers, of identifying similarities and differences and attempting to match one's speech more closely to that of others. While this process could involve techniques such as visual feedback, motor practice and perceptual discrimination, these activities would be embedded in a context of social emulation; the goal being to sound like a peer, rather than to achieve a particular motor or acoustic goal. Further research on the underlying motor and perceptual capacities for speech movements and acoustic signals will be needed to substantiate these suggestions, however.

Acknowledgements

We wish to express our appreciation to Prof. Frank Sansone for his assistance in utilizing the CSL program for these analyses, to Prof. David Snow for his guidance in the use of the accent range measure, to Jeffrey Weiing for his help in refining our analysis procedures, as well as to Elizabeth Schoen and Carolyn Gosse for their help in establishing reliability and preparing data for analysis. We also extend our gratitude to John Bianchi for his assistance with data formatting and analysis. Preparation of this paper was supported by Research National Institute of Mental Health (NIMH) Grant P01-03008; by the STAART Center grant U54 MH66494 funded by the National Institute on Deafness and Other Communication Disorders (NIDCD), the National Institute of Environmental Health Sciences (NIEHS), the National Institute of Child Health and Human Development (NICHD), and the National Institute of Neurological Disorders and Stroke (NINDS); by NIDCD MidCareer Development grant K24 HD045576 awarded to Dr. Paul; as well as by the National Alliance for Autism Research.

REFERENCES

- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. Vol. 4th. Washington, DC: Author; 1994.
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders – TR. Vol. 4th. Washington, DC: Author; 2000.
- Baltaxe C. Use of contrastive stress in normal, aphasic, and autistic children. *Journal of Speech and Hearing Research* 1984;24:97–105. [PubMed: 6201678]
- Baltaxe C, Guthrie D. The use of primary sentence stress by normal, aphasic, and autistic children. *Journal of Autism and Developmental Disorders* 1987;17:255–271. [PubMed: 2440848]
- Baltaxe, C.; Simmons, J. Communication deficits in the adolescent with autism, schizophrenia, and language-learning disabilities. In: Layton, TL., editor. *Language and treatment of autistic and developmentally disordered children*. Springfield, IL, England: Charles C Thomas; 1987. p. 155-186.
- Baltaxe, C.; Simmons, J. A comparison of language issues in high-functioning autism and related disorders with onset in children and adolescence. In: Schopler, E.; Mesibov, G., editors. *High-functioning individuals with autism*. New York: Plenum Press; 1992. p. 210-225.
- Baron-Cohen S, Staunton R. Do children with autism acquire the phonology of their peers? An examination of group identification through the window of bilingualism. *First Language* 1995;14:241–248.
- Burns, EM.; Ward, WD. Intervals, scales, and tuning. In: Deutsch, D., editor. *The psychology of music*. New York: Cambridge University Press; 1982. p. 241-269.
- Chafe, W. *Meaning and structure of language*. Chicago: University of Chicago Press; 1970.
- Couper-Kuhlen, E. *An introduction to English prosody*. *Forschung & Studium Anglistik 1*, Tübingen: Max Niemeyer and London: Edward Arnold; 1986.
- Crystal, D. *Prosodic systems and intonation in English*. The Hague: Mouton; 1969.
- Fay, W.; Schuler, AL. *Emerging language in autistic children*. Baltimore: University Park Press; 1980.
- Fay WH. On the basis of autistic echolalia. *Journal of Communication Disorders* 1969;2(1):38–47.
- Fine J, Bartolucci G, Ginsberg G, Szatmari P. The use of intonation to communicate in pervasive developmental disorders. *Journal of Child Psychology and Psychiatry* 1991;32:771–782. [PubMed: 1918227]
- Foreman C. The use of contrastive focus by high-functioning children with autism. *Dissertation Abstracts International* 2002;62:3759A. UMI No.DA3032821

- Gibbon, F.; McCann, J.; Peppe, S.; O'Hare, A.; Rutherford, M. Articulation disorders in children with high functioning autism. Paper presented at the World Congress of the International Association of Logopedics and Phoniatrics; Brisbane, Australia. 2004 Sep.
- Halliday, M. *Learning how to mean: Explorations in the development of language*. NY: Arnold; 1975.
- Kanner L. Autistic disturbances of affective contact. *Nervous Child* 1943;2:217–250.
- Kent, R.; Read, C. *The acoustic analysis of speech*. San Diego, CA: Singular Publishing Group; 1992.
- Klin A, Lang J, Cicchetti DV, Volkmar FR. Brief report: Interrater reliability of clinical diagnosis and DSM-IV criteria for autistic disorder: Results of the DSM-IV Autism Field Trial. *Journal of Autism & Developmental Disorders* 2000;30(2):163–167. [PubMed: 10832781]
- Klin, A.; Volkmar, FR. The pervasive developmental disorders: Nosology and profiles of development. In: Luthar, SS.; Burack, JA., et al., editors. *Developmental psychopathology: Perspectives on adjustment, risk, and disorder*. New York, NY: Cambridge University Press; 1997. p. 208-226.
- Klin, A.; Volkmar, FR.; Sparrow, SS., editors. *Asperger syndrome*. New York: Guilford Press; 2000.
- Koike K, Asp CW. Tennessee Test of Rhythm and Intonation Patterns. *Journal of Speech and Hearing Disorders* 1981;46:81–87. [PubMed: 7206683]
- Lehiste, I. *Suprasegmentals*. Cambridge, MA: MIT Press; 1970.
- Lord C, Risi V, Lambrecht L, Cook EH Jr, Leventhal BL, DiLavore PC, Pickles A, Rutter M. The Autism Diagnostic Observation Schedule--Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism & Developmental Disorders* 2000;30(3):205–223. [PubMed: 11055457]
- Lord C, Rutter M, LeCouteur A. Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders* 1994;24(5):659–685. [PubMed: 7814313]
- McCaleb P, Prizant B. Encoding of new versus old information by autistic children. *Journal of Speech and Hearing Disorders* 1985;50:230–240. [PubMed: 4021450]
- McCann J, Peppe S. Prosody in autism spectrum disorders: A critical review. *International Journal of Language & Communication Disorders* 2003;38(4):325–350. [PubMed: 14578051]
- Merewether FC, Alpert M. The components and neuroanatomic bases of prosody. *Journal of Communication Disorders* 1990;23(4–5):325–336. [PubMed: 2246386]
- National Institute of Deafness and Communication Disorders. (n.d.). Statistics on voice, speech, and language. Retrieved January 6, 2005 from the National Institute of Deafness and Communication Disorders Web site:<http://www.nidcd.nih.gov/health/statistics/vsl.asp>
- Ornitz EM, Ritvo ER. The syndrome of autism: A critical review. *American Journal of Psychiatry* 1976;133(6):609–621. [PubMed: 58560]
- Panagos JM, Prelock PA. Prosodic analysis of child speech. *Topics in Language Disorders* 1997;17:1–10.
- Paul, R. Communication in autism. In: Cohen, D.; Donnellan, A., editors. *Handbook of autism and pervasive developmental disorders*. New York: John Wiley & Sons; 1987.
- Paul R, Augustyn A, Klin A, Volkmar F. Perception and production of prosody by speakers with autism spectrum disorders. *Journal of Autism and Developmental Disorders* 2005;35In press
- Paul R, Shriberg L, McSweeney J, Cicchetti D, Klin A, Volkmar F. Relations between prosodic performance and communication and socialization ratings in high functioning speakers with autism spectrum disorders. *Journal of Autism and Developmental Disorders* 2005;35In Press
- Pronovost W, Wakstein MP, Wakstein DJ. A longitudinal study of the speech behaviour and comprehension of fourteen children diagnosed atypical or autistic. *Exceptional Children* 1966;33:19–26. [PubMed: 5981255]
- Quirk, R.; Greenbaum, S.; Leech, G.; Svartvik, J. *A comprehensive grammar of the English language*. NY: Longman; 1990.
- Ramberg C, Ehlers S, Nyden A, Johansson M, Gillberg C. Language and pragmatic functions in school-age children on the autism spectrum. *European Journal of Disorders of Communication* 1996;31:387–414. [PubMed: 9059572]
- Rutter M, Lockyer L. A 5 to 15 year follow-up study of infantile psychosis. I: Description of sample. *British Journal of Psychiatry* 1967;113:1169–1182. [PubMed: 6075451]

- Schwartz RG, Petinou K, Goffman L, Lazowski G, Cartusciello C. Young children's production of syllable stress: An acoustic analysis. *Journal of the Acoustical Society of America* 1996;99(5):3192–3200. [PubMed: 8642125]
- Semel, EM.; Wiig, EH.; Secord, W. *Clinical evaluation of language fundamentals - 3*. San Antonio, TX: The Psychological Corporation; 1995.
- Shriberg, L. 'Onions' and 'orchids' in phonological intervention .Workshop presented to the Northwest Speech-Language-Hearing Association Regional Convention. Seattle, WA: 1987 Oct.
- Shriberg, LD.; Kwiatkowski, J.; Rasmussen, C. *Prosody-Voice Screening Profile (PVSP): Scoring forms and training materials*. Tuscon AZ: Communication Skill Builders; 1990.
- Shriberg LD, Paul R, McSweeny JL, Klin A, Cohen DJ, Volkmar FR. Speech and prosody characteristics of adolescents and adults with high-functioning autism and Asperger syndrome. *Journal of Speech, Language, and Hearing Research* 2001;44:1097–1115.
- Solon L. Contrastive stress and children's interpretation of pronouns. *Journal of Speech and Hearing Research* 1980;23:688–698. [PubMed: 7421166]
- Snow D. Imitation of intonation contours by children with normal and disordered language development. *Clinical Linguistics and Phonetics* 2001a;15:567–584.
- Sparrow, S.; Balla, D.; Cicchetti, D. *The Vineland Adaptive Behavior Scales – Interview edition, survey form manual*. Circle Pines, MN: American Guidance Service; 1984.
- Stephens, K.; Nickerson, R.; Rollins, A. Suprasegmental and postural aspects of speech production and their effect on articulatory skills and intelligibility. In: Hochberg, I.; Levitt, H.; Osberger, M., editors. *Speech of the hearing impaired: Research, training and personnel preparation*. Baltimore: University Park Press; 1983. p. 35-51.
- Tager-Flusberg H. On the nature of linguistic functioning in early infantile autism. *Journal of Autism Developmental Disorders* 1981;11(1):45–56.
- Tager-Flusberg H. "Once upon a ribbit": Stories narrated by autistic children. *British Journal of Developmental Psychology* 1995;13(1):45–59.
- Wechsler, D. *Wechsler Intelligence Scale for Children*. Vol. 3rd. San Antonio, TX: The Psychological Corporation; 1992.
- Wechsler, D. *Wechsler Adult Intelligence Scale*. Vol. 3rd ed.. San Antonio, TX: The Psychological Corporation; 1997.

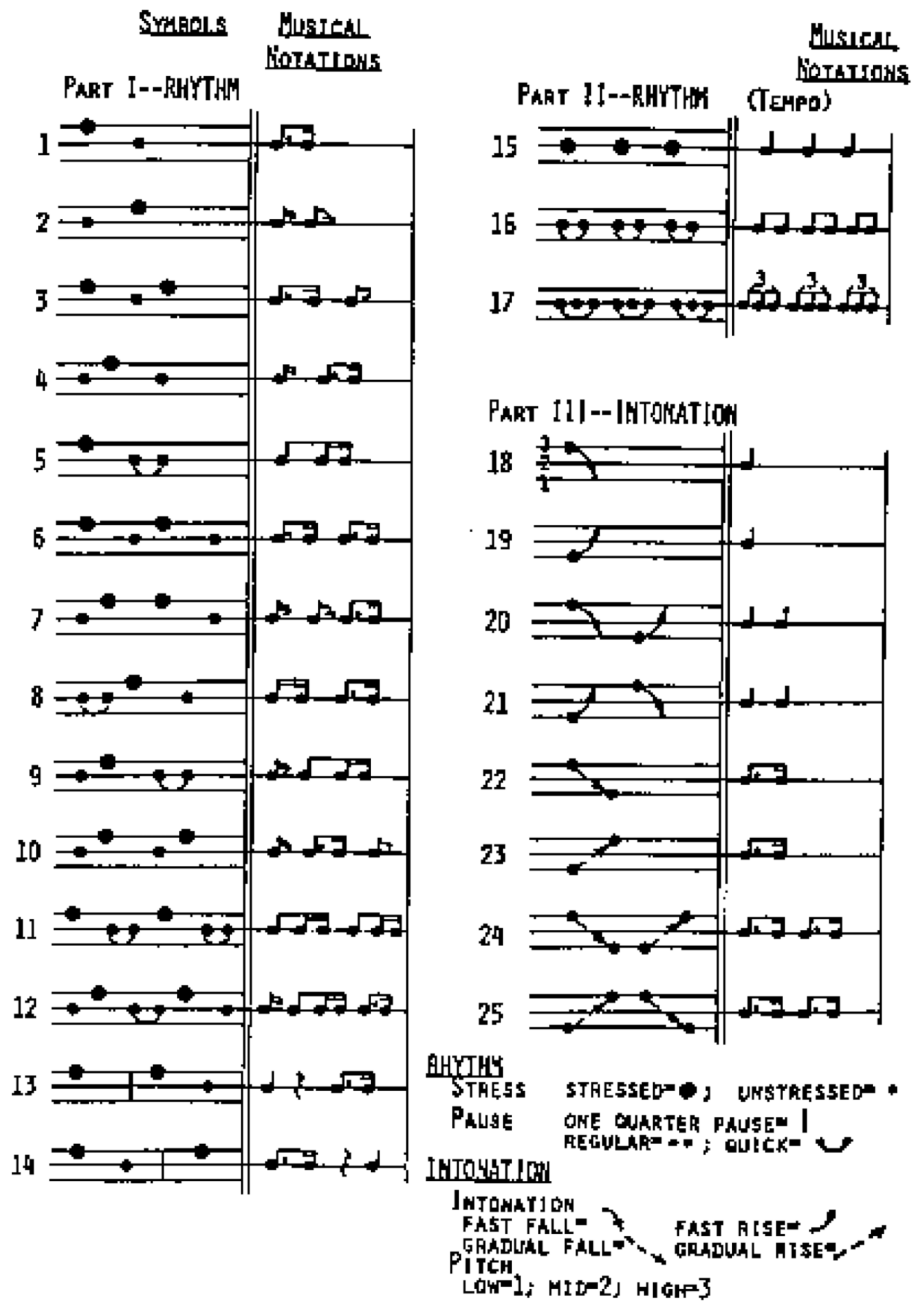


Figure 1.
Tennessee Test of Rhythm and Intonation Patterns (Koike & Asp, 1981).

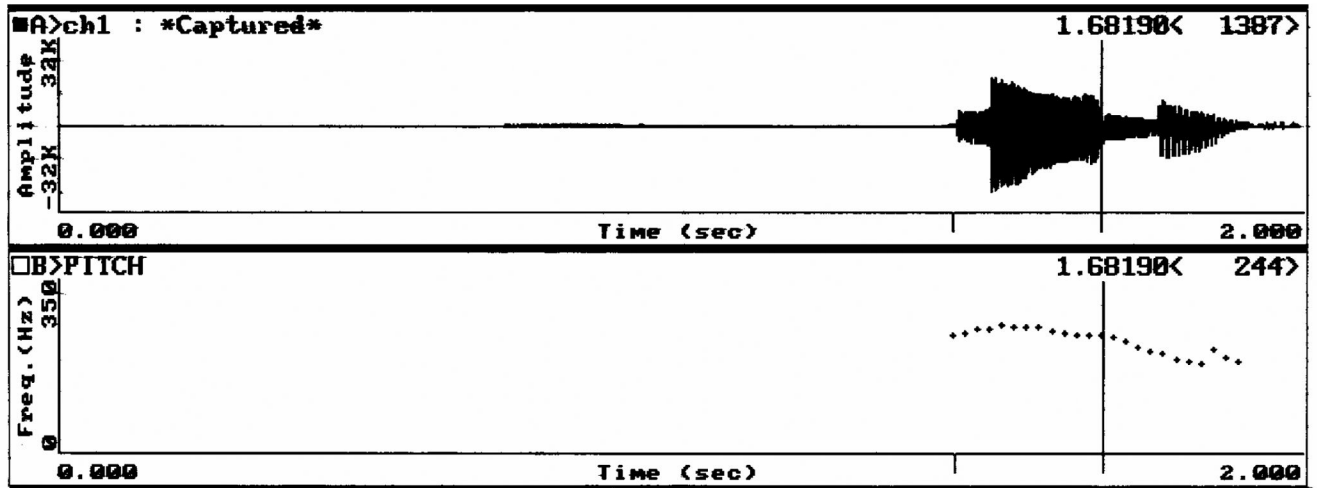


Figure 2.
Time waveform and frequency contour as depicted by the CSL program (Kay Elemetrics).

Duration in Nonsense Syllables

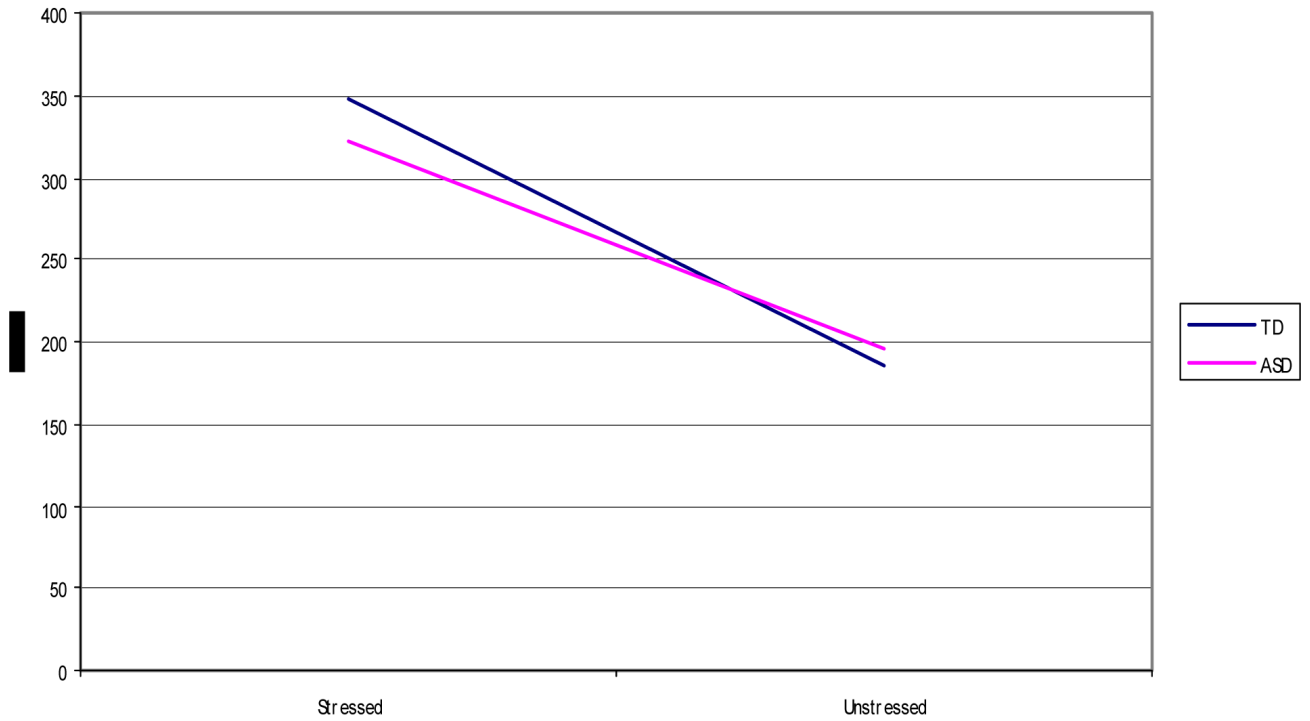


Figure 3.
Duration (milliseconds) in Syllables for Two Diagnostic Groups.

Accent Range in Nonsense Syllables

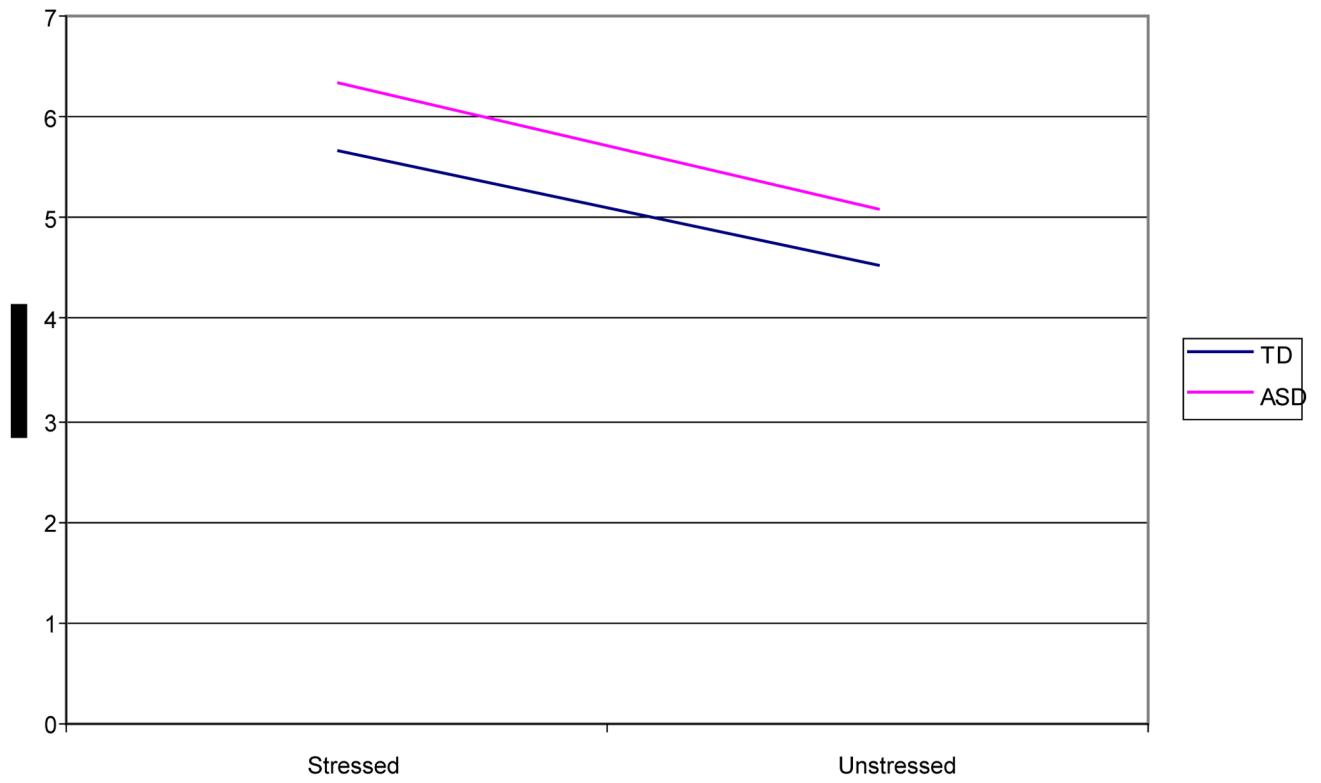


Figure 4. Accent range (semitones) in Syllables for Two Diagnostic Groups.

Table 1

	Mean (s.d.) for ASD Group (n=46)	Mean (s.d.) for HFA Ss (n=22)	Mean (s.d.) for AS Ss (n=19)	Mean (s.d.) for PDD-NOS Ss (n=5)	Significant Difference Among ASD groups?
Age	13.2 (4.4)	13.7 (5.5)	12.4 (2.9)	13.2 (4.5)	NSD
Verbal IQ ^a	103.7 (21.8)	94.8 (18.5)	115.4 (21.6)	98.4 (18.4)	F= 5.7; p< .007
Performance IQ ^a	95.0 (20.5)	90.2 (16.7)	100.2 (24.4)	96.4 (17.4)	NSD
ADOS ^b Communication Algorithm Score	4.4 (1.3)	5.1 (1.2)	4.1 (1.0)	3.0 (0.7)	F = 8.8; p< .001
ADOS ^b Social Algorithm Score	9.7 (2.6)	11.0 (2.2)	9.1 (2.1)	6.6 (2.2)	F=9.5; p<.001
Clinical Evaluation of Language Fundamentals ^c Receptive Standard Score	101.8 (21.2)	95.1 (17.5)	111.3 (23.6)	97.4 (18.4)	NSD
Clinical Evaluation of Language Fundamentals ^c Expressive Standard Score	99.7 (21.5)	92.6 (18.5)	110.1 (23.7)	93.2 (8.7)	F= 4.0; p<.03
Vineland ^d Communication Standard Score	70.8 (17.3)	67.8 (16.8)	75.7 (17.7)	65.4 (16.6)	NSD
Vineland ^d Socialization Standard Score	50.6 (13.5)	51.9 (14.0)	49.9 (12.3)	47.6 (17.5)	NSD

^aWechsler Intelligence Scale for Children, 3rd ed. (WISC; Wechsler, 1992) or Wechsler Intelligence Scale for Adults, 3rd ed. (WAIS; Wechsler, 1997), depending on subject's age

^bAutism Diagnostic Observation Scale-Module 3 or 4 depending upon subject's developmental level (Lord, et al., 2000)

^cSemel, Wiig & Secord, 1989

^dVineland Adaptive Behavior Scale (Sparrow, Balla, & Cicchetti, 1984)

Table 2

Mean (and s.d.) % correct syllable productions in four diagnostic groups

	% correct	S.D.	Range	Pair-wise comparison differences from TD; p<
Stressed				
HFA	93.7	7.7	71–100	.009*
AS	97.1	6.2	79–100	.287
PDD-NOS	87.8	14.9	66–100	.001*
TD	99.4	1.2	96–100	
Unstressed				
HFA	90.9	12.7	52–100	.001*
AS	97.6	4.2	85–100	.510
PDD-NOS	96.1	5.1	89–100	.430
TD	99.3	1.7	94–100	

Table 3

Mean (and s.d.) acoustic data on syllable productions in four diagnostic groups

Dx group	Acoustic data			
	Mean (and s.d.) Duration (msec)		Mean (and s.d.) Accent Range (semitones)	
	Stressed	Unstressed	Stressed	Unstressed
HFA	332 (59)	205 (52)	6.70 (2.87)	5.50 (1.95)
AS	328 (33)	192 (23)	6.11(2.62)	4.25 (1.81)
PDD-NOS	281 (46)	184 (41)	5.20 (1.28)	4.91 (2.00)
TD	346 (43)	186 (23)	5.69 (2.07)	4.52 (1.77)