Acoustic Differences In The Imitation Of Prosodic Patterns In Children With Autism Spectrum Disorders

Joshua John Diehl[1] and **Rhea Paul**[2]

Joshua John Diehl: joshua.diehl@nd.edu; Rhea Paul: rhea.paul@yale.edu

[1]Department of Psychology, 118A Haggar Hall, University of Notre Dame, IN, USA, 46556

[2]Yale Child Study Center, 230 South Frontage Rd., New Haven, CT, USA 06520

Abstract

In research, it has been difficult to characterize the prosodic production differences that have been observed clinically in Autism Spectrum Disorders (ASD). Moreover, the nature of these differences has been particularly hard to identify. This study examined one possible contributor to these perceived differences: motor planning. We examined the ability of children and adolescents with ASD to imitate prosodic patterns in comparison to a group with learning disabilities (LD) and a typically-developing (TD) comparison group. Overall, we found that both the ASD and LD groups were significantly worse at perceiving and imitating prosodic patterns than the TD comparison group. Similar to previous studies using non-imitative speech, participants with ASD showed a significantly longer duration of utterances than the two comparison groups when attempting to imitate an intonation pattern. The implications of differences in duration of utterances are discussed. This study also highlights the importance of using clinical comparison groups in studies of language performance in individuals with ASD.

Keywords

Autism; Imitation; Prosody; Acoustic; Communication; Pragmatics

1. Introduction

Differences in prosody production are one of the most common clinical features of Autism Spectrum Disorders (ASD; Edelson & Diehl, in press; McCann & Peppé, 2003; Paul, Augustyn, Klin, & Volkmar, 2005), although a clear and consistent characterization of the identifying features of this deficit has been somewhat elusive to researchers (Diehl & Berkovits, 2010; McCann & Peppé, 2003; Peppé, 2009). This is striking, considering that prosodic production differences are among the earliest characteristics of the disorder to appear (Oller et al., 2010; Schoen, Paul, & Chawarska, 2011; Werner, Dawson, & Osterling, 2000; Wetherby et al., 2004), and are present at all levels of ability, including Asperger syndrome (Shriberg et al., 2001) and high-functioning autism (Diehl, Watson, Bennetto, McDonough, & Gunlogson, 2009; Peppé, McCann, Gibbon, O'Hare, & Rutherford, 2007).

Corresponding Author: Joshua John Diehl, Ph.D., Department of Psychology, 118 Haggar Hall, University of Notre Dame, Notre Dame, IN 46556. Phone: (574) 631-1371. joshua.diehl@nd.edu. Fax: (574) 631-8883.

There exists a need for a better characterization of these production differences, and an understanding of the nature of the prosodic patterns observed in this population.

1.1 Prosody Production in ASD

Prosody is a broad category that refers to the patterns of rhythm, intonation (or melodic pattern), and stress (or emphasis) of the voice (Edelson & Diehl, in press). Prosodic patterns can be expressed in a single word or across an entire utterance or conversation. Functions of prosody include the communication of linguistic, pragmatic, and affective aspects of an utterance.

A number of studies have examined prosody production across multiple domains of function within a single sample. There has been some indication that individuals with ASD have trouble producing prosody at the phrasal level (e.g., Paul et al., 2005; Peppé et al., 2007) and at the lexical level (e.g., Paul et al., 2005), although the effects were surprisingly inconsistent and small, and did not seem to fully capture the extent of prosody production deficits in this population (Diehl & Berkovits, 2010). Group performance across different functions has been especially inconsistent in research, but on an individual level most individuals with ASD seem to have difficulty in at least one aspect of prosodic ability (McCann, Peppé, Gibbon, O'Hare, & Rutherford, 2007). What was more consistent in these (and other) studies was that participants with ASD, especially high-functioning individuals, performed very well on many of the tasks.

1.2 Acoustic Analysis of Prosody Production in ASD

Recently, researchers have begun using acoustic analysis of speech in order to measure more subtle prosodic differences. Acoustic analysis provides an objective measure of speech performance, and it affords the researcher the ability to examine details of elements of production rather than forcing an assignment of a production to a binary category (i.e., correct or incorrect). Until recently, the acoustic analysis of speech has been quite difficult, albeit important (Green & Tobin, 2009; Peppé, 2009), which is one reason why research in this area has lagged behind other areas of linguistics, such as syntax and semantics, even for typical development.

Studies have acoustically analyzed echolalic, imitative, and spontaneous speech in conversation and narratives in children, adolescents, and adults with ASD. Several of these studies have found important prosodic differences in the variance of fundamental frequency (f0), duration of syllables, the use of prosodic contours (Baltaxe, 1984; Diehl et al., 2009; Esposito & Venuti, 2008; Fosnot & Jun, 1999; Green & Tobin, 2009; Grossman, Bemis, Plesa Skwerer, & Tager-Flusberg, 2010; Loca & Wootton, 1995; Paccia & Curcio, 1982), and the coordination of multiple prosodic cues such as pitch, duration, and amplitude (Van Santen, Prud'hommeaux, Black, & Mitchell, 2010). Diehl and Paul (in press) found that several acoustic differences in prosody were present in the speech of individuals with ASD even when the prosodic cues were perceived as functionally correct (i.e., responses were judged by rater), suggesting that prosodic production differences involve more than just differences in the functional use of prosody. Several tools are being developed that process the acoustic signal for the purpose of early (and automated) detection and diagnosis of ASD (Oller et al., 2010; Van Santen et al., 2010; Warren et al., 2011).

1.3 The Nature of Prosodic Production Differences in ASD

It is not surprising that even less is known about the origin of production differences than how they are characterized. Given the heterogeneity of the autism spectrum as a whole and the array of clinically-observed prosodic patterns in this population, it is unlikely that there will be a single explanation for the etiology of prosody production deficits. Several theories

have been suggested that can account for some (but not all) of the observed differences. One proposal has been that a deficit in understanding and communicating affect and/or other internal states might underlie these differences (e.g., Rutherford, Baron-Cohen, & Wheelwright, 2002), given the extent of the pragmatic deficits across the spectrum (Tager-Flusberg, Paul, & Lord, 2005; Young, Diehl, Morris, Hyman, & Bennetto, 2005). Although deficits in affective/mental state processing and production have been found (Kleinman, Marciano, & Ault, 2001; Rutherford et al., 2002), they are not always present (Grossman et al., 2010), and do not fully account for the differences found in other domains, such as syntactic structure (Diehl, Bennetto, Watson, Gunlogson, & McDonough, 2008).

More broadly, Shriberg and colleagues (2011) have suggested that speakers with ASD have difficulty "tuning up" their productions to emulate models provided by other speakers, so that subtle differences that are detectable without necessarily affecting category boundaries, are present. That is, some of the prosody production differences in ASD could be attributed to a difficulty in attunement to ambient conventions for community-acceptable production. In essence, individuals with ASD might possess a dearth of social motivation to "talk just like" other speakers in the community (Paul, Bianchi, Augustyn, Klin, & Volkmar, 2008). This might explain why individuals with ASD fail to acquire a community-appropriate accent or dialect (Baron-Cohen & Staunton, 1994) and show persistent distortions in speech sounds that are "outgrown" by typical speakers (Shriberg et al., 2001).

One possible and underexplored contributor to prosodic differences is the motor planning and execution aspect of prosodic production. There is extant literature that individuals with ASD have imitation deficits that are present very early in development (Rogers & Williams, 2006). Despite these findings, surprisingly little has been done to investigate basic imitation abilities for prosody production; although it has been argued that the pattern of deficits seen in ASD are not the same as the typical pattern observed in individuals with motor planning deficits (Shriberg et al., 2011). A series of studies used the Profiling Elements of Prosodic Systems – Children test (PEPS-C; Peppé & McCann, 2003) to look at imitation of prosodic patterns (Järvinen-Pasley, Peppé, King-Smith, & Heaton, 2008; McCann et al., 2007; Peppé et al., 2007; Peppé, McCann, Gibbon, O'Hare, & Rutherford, 2006). These studies found consistent differences between children with ASD, typically-developing peers, and adults imitating prosodic patterns in words and sentences. Children with ASD were much less likely to be judged as correctly imitating the prosodic patterns than comparison groups. These studies, however, relied on correct versus incorrect responses, rather than using acoustic analyses, which made it difficult to understand where these prosodic differences were occurring (e.g., duration, intensity).

Two studies have used acoustic analysis of speech to examine differences in prosody imitation. In one study, Paul and colleagues examined imitated speech in 44 children and adults with ASD (ages 7–28) in comparison to 20 individuals with typical development (Paul et al., 2008). Participants listened to and imitated stressed and unstressed nonsense syllables. Paul and colleagues found significant differences in the duration of vocalizations, despite the fact that they predicted no differences. Specifically, participants with ASD showed significantly less difference in duration between stressed and unstressed syllables than the typically developing comparison group. They also found that pitch range tended to be larger for participants with ASD for both stressed and unstressed syllables, although this difference failed to reach statistical significance.

In another study, Van Santen and colleagues used automatic analysis of speech to examine prosody production in children with ASD (ages 4–8) in comparison to typically-developing peers. Similar to Paul et al. (2008), participants completed a vocal imitation paradigm in which they had to repeat stress patterns of two-syllable nonsense words. In a second task,

they imitated entire sentences in which a number of words were stressed. Participants also gave samples of non-imitative speech in a picture description paradigm. The study found that group differences in imitation were present in one of the two (single word) imitation tasks, but the differences were smaller than in the spontaneous production task. Interestingly, in the other imitation task, individuals with ASD were judged by examiners to have worse imitation ($p<.002$), but acoustic measures indicated that they were better at imitation. The authors argue that experimenters or judges have a tendency to pay too much attention to average pitch, and do not pay attention to other differences in production between groups, such as duration of utterance.

In sum, there is some evidence that individuals with ASD show differences in imitation of prosodic patterns (McCann et al., 2007; Paul et al., 2008; Peppé et al., 2007; Peppé et al., 2006), although the imitation deficits do not seem to explain all of the variance in prosodic patterns in this population (Van Santen et al., 2010). Moreover, other studies suggest that there are no motor planning issues, and patterns of production can be better explained by social-cognitive factors (Shriberg et al., 2011). Importantly, only two of these studies used acoustic analysis of imitative speech, and neither of the studies included a clinical comparison group. This latter point is important because there is evidence that speakers with other language-related developmental disabilities also exhibit difficulties in the processing and production of prosody (Catterall, Howard, Stojanovik, Szczerbinski, & Wells, 2006; Marshall, Harcourt-Brown, Ramus, & van der Lely, 2009; Stojanovik, Setter, & Ewijk, 2007; Wells & Peppé, 2003), although these are less commonly reported, and the investigation of their characteristics is less detailed. Also, to date there are no acoustic studies on imitation using the PEPS-C, a test which consistently produced imitative deficits.

1.4 Purpose of the Study

The purpose of the present study was to investigate the ability of children and adolescents with ASD to imitate prosodic patterns in single words and phrases using perceptual judgment of responses and acoustic analysis of speech. We compared the performance of the ASD group to a group with learning disabilities (LD) and a typically-developing (TD) comparison group using the PEPS-C, a test designed to examine prosody perception, production, and imitation in typical and atypical populations. The PEPS-C is useful for this study and acoustical analysis in particular because it has been used in several studies to examine prosodic performance in ASD as well as in language impairment (Wells & Peppé, 2003), and the production content is standardized across participants, which allows for comparisons of multiple acoustic characteristics of prosody such as pitch range, duration, and intensity. Based on findings from a series of studies indicating poor performance by individuals with ASD on the imitation portions of this test, we predicted that participants with ASD would perform worse than the two comparison groups on the percentage of responses judged to be correct by the examiner. Additionally, we predicted that the ASD group would show longer duration of utterances and larger pitch range than the two comparison groups, similar to a previous study using acoustic analysis of elicited (rather than imitated) prosody on this test (Diehl & Paul, in press), and we predicted the differences to be smaller than seen for spontaneous speech. Although Diehl & Paul (in press) found acoustic differences between the LD and TD groups, we predicted that these differences would disappear in an imitation task.

2. Methods

2.1 Participants

2.1.1 Autism Spectrum Disorders—Participants in this group included 24 children and adolescents with ASD between the ages of 8 and 16. Participants were recruited for this

study from an existing database of children/adolescents who had participated in research and/or received clinical services at the Yale Child Study Center during the five years previous to the study and had given permission to be contacted for future research studies. Participants were diagnosed based on the Autism Diagnostic Interview-Revised (ADI-R; Rutter, Le Couteur, & Lord, 2003) and the Autism Diagnostic Observation Schedule-Generic (ADOS-G; Lord et al., 2000). Diagnoses were confirmed independently by two experienced clinicians, who determined that each participant currently met DSM-IV-TR (APA, 2000) criteria for one of the three ASDs (Autistic Disorder, Asperger's Disorder, or Pervasive Developmental Disorder, Not Otherwise Specified). Inter-rater reliability for diagnostic assignment between these two clinicians was high (kappa range = .80–.95) in related research projects (Klin, Lang, Cicchetti, & Volkmar, 2000). Participants with ASD received the Clinical Evaluation of Language Fundamentals – Fourth Edition (CELF-IV; Semel, Wiig, & Secord, 2003) in order to determine the current level of language functioning. Exclusion criteria included any uncorrected sensory disorders (vision, hearing) or known neurological conditions (see Table 1 for descriptive characteristics of the sample).

2.1.2 Learning disabilities comparison group—The learning disabilities (LD) comparison group was comprised of 16 children and adolescents (ages of 9–16) who were recruited through a community speech-language pathologist. Participants in the LD group all had a learning disability that was clinically diagnosed by a speech-language pathologist, and all were at the appropriate grade level for their chronological age. The group with LD had a number of learning disability diagnoses, many (but not all) of which were language based. All were screened via clinical interview and were found not to exhibit ASD. Additionally, family history showed no evidence of a first degree relative with an ASD diagnosis. Participants in the LD group received a brief cognitive and language evaluation. The ASD and LD groups were matched on chronological age, and had similar nonverbal IQ's, CELF-IV Core Language, and CELF-IV Receptive Language Index scores. Consistent with the fact that many participants in the group with LD had a language-based LD, the ASD group tended to have higher scores on CELF-IV indices (Expressive Language, in particular).

2.1.3 Typically-developing comparison group—The typically-developing (TD) group included 22 children (ages 8–17) who were recruited from the community. Participants in this group all had typical development as reported by their parents. All TD participants had no first degree relatives with an ASD, no previous history of clinical diagnosis or special educational services, and were in the expected grade in school for their age. Participants in the TD group were matched on chronological age to the other two groups. All three groups were identical to the sample used in Diehl and Paul (in press).

2.2 Procedures

2.2.1 Experimental setup—Participants sat at a table in front of a Dell Inspiron 3900 computer that contained the PEPS-C (Peppé & McCann, 2003) program. Participants wore a Shure SM10A Professional Unidirectional Head-mounted Dynamic Microphone, which was connected to a TASCAM US-122 USB Audio/MIDI Interface. The TASCAM connected directly into the Dell computer. Recordings had a sampling rate of 44.1 kHz.

2.2.2 Measures and Stimuli—All three groups were administered the PEPS-C, a test designed to assess the perception and production of prosody in children and adolescents between the ages of 4 and 16 (Wells & Peppé, 2004). The PEPS-C is perhaps the most widely used standard measure of prosodic performance in the literature (Peppé, 2009). The PEPS-C has been used to gather normative data from typically-developing children (Peppé & McCann, 2003). Additionally, several studies have used the test administrators' perceptual judgment of participant to investigate the ability to produce and perceive prosody in children

with a range of disabilities (Catterall et al., 2006; Diehl & Paul, in press; Marshall et al., 2009; McCann et al., 2007; Peppé et al., 2007; Peppé & McCann, 2003; Peppé et al., 2006; Stojanovik et al., 2007; Wells & Peppé, 2003).

The PEPS-C contains 12 subtests divided into six categories (affect, chunking, turn-end type, focus, intonation, prosody). For each of the categories, there is an "Input" (perception) and "Output" (production) subtest. Two of the categories (intonation, prosody) are "Form" categories, and are comprised of simple pitch/melody discrimination and productive imitation; whereas, the other four categories (affect, chunking, turn-end type, and focus) are "Function" categories that are measures of participants' ability to understand and produce prosody in a way that communicates a specific function, such as an affective state. Acoustic and behavioral response data from this sample from the Function tasks are reported in a separate paper (Diehl & Paul, in press).

For each of the subtests, and consistent with the PEPS-C instructions, the experimenter demonstrated the correct answer on two training trials in order to explain the task to the participant. The participant then completed two additional practice items, during which the experimenter would correct the participant for incorrect responses in order to ensure that the participant understood the task. Each subtest consisted of 16 experimental trials. For the two Form Input subtests, we collected behavioral data. Participants indicated their response by clicking on one of two possible response choices, and the PEPS-C computer program automatically coded the response as correct or incorrect. For the Form Output subtests (see Table 2), responses were judged as correct or incorrect by a trained examiner. For the two Output subtests that involved examiner judgment, responses from a randomly selected 10% sample of participants were independently scored by a second examiner. The average point-to-point reliability for correct/incorrect judgments on each subtest across participants ranged from .84 to .96, with an average agreement of .88 across subtests. It should be noted that one test is called "Intonation," and the other is called "Prosody," although both tests involve prosody. The key prosodic features in the Intonation Output task are intonation patterns; whereas, the important prosodic features in Prosody Output are rhythm and pauses.

2.2.3 Acoustic analysis—All output responses were audio-recorded, and vocalizations from the critical trials on all of the Form Output subtests were examined for the following acoustic characteristics of prosody: utterance duration, utterance intensity, average f0 across the entire utterance, the difference between maximum and minimum f0, or f0 (accent) range, and the standard deviation (SD) of f0 (as measured in Diehl et al., 2008; Diehl & Paul, in press). Duration is a simple measure of the speed and rhythm of speech, and intensity is a measure of the loudness or quietness of an utterance. Average f0 is a measure of how high or deep a person's voice is, and it is an important measure to include because prosodic patterns can vary in other acoustic characteristics based on the average f0 (e.g., higher average f0 generally translates into higher pitch/accent range). The f0 range and SD of f0 are standard measures of the pitch range of speech, and these data can indicate whether or not prosodic patterns are larger (i.e., more exaggerated) or smaller (e.g., less variable, more monotone) than would be expected. We used PRAAT, a program for speech analysis and synthesis (Boersma & Weenink, 2009), for acoustic analyses. Sound files were automatically divided into individual trials by the computer. In PRAAT, text grids were created to delineate the beginning and end of each vocalization. PRAAT scripts were created to extract automatically the acoustic characteristics for each trial.

3. Results

3.1 Data Analysis Plan

All analyses were initially conducted using one-way Analyses of Variance (ANOVA) with group membership (ASD, LD, TD) as the independent variable and either behavioral response accuracy or one of the five acoustic measures as the dependent variables. First, an omnibus one-way ANOVA was conducted including all three groups. Because of the large number of measurements, we only conducted paired diagnostic group comparisons (i.e., ASD vs. LD, ASD vs. TD, LD vs. TD) if the omnibus F was significant, $p<.05$, in order to reduce the likelihood of Type I error. Effect sizes were calculated as partial eta squared ($\eta2_{partial}$) which refers to the proportion of variance attributable to a given effect, after partialling out other non-error sources of variance (Cohen, 1973). Data on behavioral measures appear in Figures 1 and 2. Means and standard errors for all acoustic measures appear in Figures 3 though 6.

We tested all comparisons for violations of the assumptions of ANOVA. For all of the behavioral analyses (both Form categories for both Input and Output subtests), there was significant heteroscedacity (Levene's test) because the typically-developing comparison group did not miss very many items on any of the subtests. To correct for this issue, Welch's (W) test was used on all statistical analyses on the behavioral data that had heterogeneity of variance (i.e., all that included the typically-developing comparison group); whereas, the standard F test was used for all other comparisons.

3.2 PEPS-C Intonation Subtests

3.2.1 Behavioral findings from Intonation Output/Input subtests—For the
Intonation Output subtest, we measured participants' ability to imitate intonation patterns of single words (see Table 2). A Welch's one-way analysis of variance (ANOVA) revealed significant group differences overall, $W(2,25.39)=9.38$, $p<.001$. The TD group performed significantly better than the group with ASD, $W(1,24.13)=13.19$, $p<.001$, and the group with LD, $W(1,15.77)=6.51$, $p<.05$, but the groups with ASD and LD did not differ in their performance, $F(1,38)=3.88$, $p=.44$, $\eta^2_{partial}=.02$. Intonation Output scores in the ASD and LD groups were not significantly correlated with general language abilities as measured by the CELF Core Language score, $r(38)=.09$, $p=.57$.

For Intonation Input, we measured participants' ability to discriminate between the intonation patterns of single words, but without the consonant or vowel sounds that would distinguish the words (see Table 2). A Welch's one-way ANOVA revealed significant group differences overall, $W(2,24.79)=3.60$, $p<.05$. The TD group performed significantly better than the group with LD, $W(1,15.16)=4.35$, $p<.05$, and marginally better than the group with ASD, $W(1,24.05)=3.12$, $p=.09$, but the groups with ASD and LD did not differ in their performance, $F(1,38)=1.47$, $p=.23$, $\eta^2_{partial}=.04$. Intonation Input scores in the ASD and LD groups were significantly correlated with general language abilities (CELF-IV Core Language, $r(38)=.33$, $p<.05$).

3.2.2 Intonation Output subtest acoustic findings—For items where participants
had to imitate the prosodic pattern of single words (PEPS-C Intonation Output), one-way ANOVAs revealed significant group differences in utterance duration, $F(2,59)=3.49$, $p<.05$, $n^2_{partial}=.11$, a trend for differences in average f0, $W(2,31.95)=2.57$, $p=.09$, but no significant differences in intensity, $W(2,35.37)=2.5$, $p=.53$, f0 range, $F(2,59)=1.04$, $p=.36$, $n^2_{partial}=.03$, or SD of f0, $W(2,33.54)=1.18$, $p=.32$. The ASD group had significantly longer utterances than the TD group, $F(1,44)=5.71$, $p<.05$, $n^2_{partial}=.12$, but there was not a significant difference in utterance duration between the ASD and LD groups,

$W(1,36.67)=2.64$, $p=.11$, or between the group with LD and the TD group, $F(1,36)=1.16$, $p=.30$, $n^2_{partial}=.03$. Post hoc analyses revealed that duration of utterance was neither correlated with CELF-IV scores overall (Core Language, $r(38)=-.07$, $p=.67$, Receptive Language $r(38)=.04$, $p=.80$, or Expressive Language, $r(38)=-.06$, $p=.73$) nor when the group with ASD was examined alone (Core, $r(22)=-.10$, $p=.65$, Receptive, $r(22)=-.04$, $p=.86$, or Expressive, $r(22)=-.12$, $p=.59$)

3.3 PEPS-C Prosody Subtests

3.3.1 Behavioral findings from Prosody Output/Input subtests—For the Prosody Output subtest, we measured participants' ability to imitate prosodic patterns of short phrases (see Table 2). A Welch's one-way ANOVA revealed significant group differences overall, $W(2,29.80)=6.63$, $p<.01$. The TD group performed significantly better than the ASD group, $W(1,31.99)=9.32$, $p<.01$, and LD group, $W(1,18.40)=5.97$, $p<.05$, but the groups with ASD and LD did not differ in their performance, $F(1,38)=.01$, $p=.93$, $\eta^2_{partial}=.001$. Prosody Output scores were significantly correlated with general language abilities (CELF Core Language, $r(38)=.41$, $p<.01$).

For the Prosody Input subtest, we measured participants' ability to discriminate between sound patterns that mimicked prosodic patterns of short phrases, but contained no consonant or vowel sounds that would allow the listener to perceive words (see Table 2). A Welch's one-way ANOVA revealed significant group differences overall, $W(2,30.43)=30.43$, $p<.05$. The TD group performed significantly better than the group with ASD, $W(1,36.25)=4.05$, $p<.05$, and the group with LD, $W(1,18.08)=7.78$, $p<.01$, but the groups with ASD and LD did not differ in their performance, $F(1,38)=2.36$, $p=.13$, $\eta^2_{partial}=.06$. Intonation Input scores were significantly correlated with general language abilities as measured by CELF Core Language, $r(38)=.44$, $p<.01$.

3.3.2 Prosody Output subtest acoustic findings—For items where participants had to imitate the prosodic pattern of short phrases (PEPS-C Form Prosody Output), one-way ANOVAs revealed a significant group difference in average f0, $(2,31.64)=3.55$, $p<.05$, but no significant differences in utterance duration, $F(2,59)=.86$, $p<.43$, $n^2_{partial}=.03$, utterance intensity, $W(2,36.92)=1.64$, $p=.21$, f0 range, $F(2,59)=.25$, $p=.78$, $n^2_{partial}=.01$, or SD of f0, $F(2,59)=.11$, $p=.90$, $n^2_{partial}=.004$. The LD group had a significantly lower average f0 than the ASD group, $F(1,38)=6.03$, $p<.05$, $n^2_{partial}=.14$, and the TD group, $W(1,34.04)=5.94$, $p<.05$, but average f0 was not significantly different between the ASD and TD groups, $W(1,33.45)=.47$, $p=.50$.

4. Discussion

In this study, we examined the ability of children and adolescents with ASD to perceive similarity between prosodic productions and to imitate prosodic patterns in single words and phrases in comparison to LD and TD comparison groups. We used perceptual judgment of responses and acoustic analysis of speech to examine differences in imitation. To elicit responses, we used the PEPS-C, a test designed to examine prosody perception, production, and imitation in typical and atypical populations. We predicted that the participants with ASD would be worse at imitating prosody than both groups, and that these differences would be observable in the acoustic analysis of speech. Our hypothesis was partially supported, as both the ASD and LD groups performed worse than the TD groups on two imitation tasks and two perceptual discrimination tasks. For the acoustic analyses, the group with ASD tended to have a longer duration of utterances, but there was a surprising lack of acoustic differences across the two subtests.

4.1 Perceptual Analyses

For the perceptual analyses, we predicted that participants with ASD would have greater difficulty imitating prosodic patterns in speech than LD or TD comparison groups. For Output tasks, scores were judged as correct/incorrect by a rater, and for Input tasks, the computer automatically scored the answers. Similar to previous studies using the PEPS-C (e.g., Peppé et al., 2007), we found consistent group differences between the ASD group and TD peers on the imitation of prosodic patterns (Intonation and Prosody Form Output Tasks) and also the discrimination of prosodic patterns (Intonation and Prosody Input Tasks). Contrary to our predictions, however, the LD group also performed significantly worse than the TD group on all of these tasks as well, and the ASD and LD groups did not differ on their performance. One possible explanation is that we did not closely match groups on CELF scores. Indeed, performance on all four PEPS-C subtests was significantly correlated with CELF-IV Core Language scores (r's ranging from .36 to .51). We did not match on this measure because we wanted to capture acoustic differences that would be typical of ASD, rather than from a very restricted subset. Still, these findings highlight the importance of considering level of general language functioning when examining aspects of pragmatic communication.

4.2 Acoustic Analyses

For the acoustic analyses, we predicted that the ASD group would show longer duration of utterances and larger pitch range than the two comparison groups; although, we predicted the differences to be smaller than seen from findings in a previous study using acoustic analysis of non-imitative prosody on this same sample (Diehl & Paul, in press). Although our previous study found acoustic differences between the LD and TD groups, we predicted that these differences would disappear in an imitation task. We found partial support for our predictions. The group with ASD had a significantly longer duration of utterances on one of the two (single word) imitation subtests, but we did not find group differences in either pitch range or most other acoustic categories.

These data are consistent with several studies in addition to Diehl and Paul (in press) that have found differences in duration of utterances for imitative and spontaneous speech in high-functioning individuals with ASD (although see Shriberg et al., 2011, for different findings). Grossman and colleagues (2010) found increased duration of stressed syllables that lead to longer total duration of words in a sentence completion task than typically developing peers. Paul and colleagues (2008) found that individuals with ASD did not use duration as a means to differentiate stressed syllables from unstressed syllables in the same way that typically-developing peers did when imitating nonsense words. Van Santen and colleagues (2010) noted that duration was one of the most important acoustic features differentiating individuals with ASD from their TD peers. The striking finding from this study was that duration was the only acoustic characteristic that differentiated the ASD group from the typically developing group.

Taken together, these findings can be interpreted to suggest that timing of speech units appears to be somewhat distorted in the connected speech of speakers with ASD. Whether these subtle timing differences are attributable to motor difficulties that limit speed of execution or transition to the next unit cannot be answered definitively by this study, but the findings do raise the question. Shriberg et al. (2011) reported not finding significantly slower rate of speech in young (4–7 year old) children with ASD, but a range of other studies do report increased durations in ASD in older children. It could be the case that as typically developing children mature, they decrease the duration of syllables with increasing motor skill, and children with ASD fail to achieve this decrease. Again, whether this is a result of motor problems, or sensitivity to typical durations in the speech community is

unclear. Studies of diadochokinetic performance in children with ASD at various ages may help to elucidate this finding.

It should be noted that this is the second study using the PEPS-C on this sample, with the first study examining acoustic characteristics of prosodic functions (Diehl & Paul, in press). Interestingly, the items in the Intonation Output subtest are nearly identical to the spontaneous speech elicited by the Affect and Turn-End Type subtests in the PEPS-C. When comparing imitative vs. spontaneous speech with this same sample (see Figure 7), participants with ASD produced durations that more closely mimicked the test item when imitating speech in contrast to items that were spontaneous productions. By comparison, the LD and TD groups actually produced imitated utterances that were longer, and less like the duration of the test items, than when these words were spontaneously produced. This could reflect the perception of children with TD and LD because these tokens were non-communicative, a rate comparable to that used in real speech was not necessary; whereas, the participants with ASD treated communicative and non-communicative material just the same way.

4.3 Limitations and Future Directions

One limitation was that we found significant differences in average f0 across the groups in the two tasks. This is a limitation, rather than a notable finding, because average f0 is an indicator of how high or deep a person's voice is, rather than a meaningful prosodic factor. Importantly, voices with a higher average f0 tend to have a wider pitch range by nature, and the relatively higher average f0 for the ASD group and the relatively lower average f0 for the LD group might have affected the findings on the two measures of pitch range. Interestingly, it is most likely that matching on average f0 would have increased pitch range in the LD group and decreased it in the ASD group, a pattern which would still have continued to be contrary to our hypothesis that individuals with ASD would have a higher pitch range.

An additional limitation is that behavioral performance on the input and output subtests for both intonation and prosody were strongly correlated with CELF Core Language, RLI, and ELI for all three groups, and because there were small differences in language scores between the ASD and LD groups, we may be seeing effects of general language level in the observed differences, rather than effects specific to prosody. However, it is noteworthy that duration in the ASD group is not at all correlated with CELF scores or performance on any of the output or input tasks. This could be taken to suggest that, although general language skill provides some explanation of accuracy of response, the acoustic character of the response is not so influenced. This observation could be tested further by including working memory and other cognitive measures to determine the cognitive factors involved in performance differences. In addition, studying larger samples that would allow the use of regression analyses would allow the testing of predictive capacities among variables such as general cognitive, language and motor measures on acoustic properties of production, such as duration.

In addition, the findings of this study strengthen the suggestion that in order to understand autistic behavior such as prosodic deficits, it is crucial to use contrast groups composed of carefully matched children with developmental disabilities who do not have an ASD. As our findings, which highlight unexpected similarities between prosodic production in ASD and LD emphasize, these comparisons are necessary in order to fully determine what aspects of behavior and development are unique to ASD, and which are shared among a range of disorders. In the area of language particularly, this careful disentangling is critical to understanding the complementary and intertwined roles played by language and social development in this and other syndromes.

4.4 Clinical Implications

Our study identified prosodic deficits even in simple imitation tasks of words and phrases, in children with both ASD and LD. For children with ASD, performance on the imitation and elicited speech tasks were markedly similar, suggesting they may not be attending to the communicative value of the non-imitated units, whereas the children with LD did seem more affected by this difference. Although we would not argue that the subtle acoustic differences observed in this study should be the focus of intervention attempts, it may be useful to work with students with ASD on perceiving, thinking about, and making changes in communicative vs. non-communicative speech production as a way to help them focus their attention on the interactive burden of communicative speech. Rather than focusing on a specific acoustic parameter, a clinician might simply give students with ASD phrases to read either to another (perhaps as instructions to play a game) or merely to practice reading aloud (or "exercising the voice") without communicative intent. Such activities might serve as a platform for discussing and calling the student's attention to how we talk differently when we really need to be understood. For students with LD, the data suggest more clinical attention be paid to the prosodic form of speech, again not necessarily to correct particular prosodic errors, but more generally to bring conscious awareness to the paralinguistic functions prosody can convey, in an effort to help focus attention, provide enhanced modeling, and practice in "tuning up" to the prosodic features of the speech community.

4.5 Conclusions

In sum, it has been difficult to characterize the prosodic production differences in ASD. Moreover, it has been difficult to understand their nature, given the heterogeneity of the autism spectrum as a whole and the array of clinically-observed prosodic patterns in this population, it is unlikely that there will be a single explanation for the etiology of prosody production deficits. This study suggests that the use of duration is an important factor and that potential causes for differences in duration, such as levels of cognition, general language and motor ability warrant further examination.

Acknowledgments

References

American Psychiatric Association. Diagnostic and statistical manual of mental disorders, text revision. 4th ed.. Washington, D.C: APA; 2000.

Baltaxe CAM. Use of contrastive stress in normal, aphasic, and autistic children. Journal of Speech and Hearing Research. 1984; 27:97–105. [PubMed: 6201678]

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. 1994; 14:241–248.

Boersma P, Weenink D. Praat: Doing phonetics by computer (version 5.1.20) [computer program]. 2009 Retrieved from http://www.praat.org/.

Catterall C, Howard S, Stojanovik V, Szczerbinski M, Wells B. Investigating prosodic ability in Williams syndrome. Clinical Linguistics and Phonetics. 2006; 20:531–538. [PubMed: 17056483]

Cohen J. Eta-squared and partial eta-squared in fixed factor ANOVA designs. Educational and Psychological Measurement. 1973; 33:107–112.

Diehl JJ, Bennetto L, Watson D, Gunlogson C, McDonough J. Resolving ambiguity: A psycholinguistic approach to understanding prosody processing in high-functioning autism. Brain and Language. 2008; 106:144–152. [PubMed: 18502497]

Diehl, JJ.; Berkovits, L. Is prosody a diagnostic and cognitive bellwether of autism spectrum disorders?. In: Harrison, A., editor. Speech disorders: Causes, treatments, and social effects. New York: Nova Science Publishers; 2010. p. 159-176.

Diehl JJ, Paul R. Acoustic and perceptual measurements of prosody production on the PEPS-C by children with autism spectrum disorders. Applied Psycholinguistics. (in press).

Diehl JJ, Watson DG, Bennetto L, McDonough J, Gunlogson C. An acoustic analysis of prosody in high-functioning autism. Applied Psycholinguistics. 2009; 30:385–404.

Edelson, L.; Diehl, JJ. Encyclopedia of Autism Spectrum Disorders. Springer: Prosody. (in press)

Elliott, CD. Differential ability scales. San Diego: Harcourt Brace Jovanovich; 1990.

Esposito G, Venuti P. How is crying perceived in children with autism spectrum disorder? Research in Autism Spectrum Disorders. 2008; 2:371–384.

Fosnot, SM.; Jun, S. Prosodic characteristics in children with stuttering or autism during reading and imitation. Paper Presented at the 14th International Congress of Phonetic Sciences; 1999.

Green H, Tobin Y. Prosodic analysis is difficult…but worth it: A study in high functioning autism. International Journal of Speech-Language Pathology. 2009; 11:308–315.

Grossman RB, Bemis RH, Plesa Skwerer D, Tager-Flusberg H. Lexical and affective prosody in children with high-functioning autism. Journal of Speech, Language, and Hearing Research. 2010; 53:778–793.

Järvinen-Pasley A, Peppé S, King-Smith G, Heaton P. The relationship between form and function level receptive prosodic abilities in autism. Journal of Autism and Developmental Disorders. 2008; 38:1328–1340. [PubMed: 18172749]

Kleinman J, Marciano PL, Ault RL. Advanced theory of mind in high-functioning adults with autism. Journal of Autism & Developmental Disorders. 2001; 31:29–36. [PubMed: 11439751]

Klin A, Lang J, Cicchetti DV, Volkmar FR. Brief report: Interrater reliability of clinical diagnosis and DSM-IV criteria for autistic disorder: Results of DSM-IV autism field trial. Journal of Autism and Developmental Disorders. 2000; 30:163–167. [PubMed: 10832781]

Loca J, Wootton T. Interactional and phonetic aspects of immediate echolalia in autism - a case study. Clinical Linguistics & Phonetics. 1995; 9:155–184.

Lord C, Risi S, Lambrecht L, Cook EH, Leventhal BL, DiLavore PC, Pickles A, Rutter M. Autism diagnostic observational schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism. Journal of Autism and Developmental Disorders. 2000; 30:205–223. [PubMed: 11055457]

Marshall CR, Harcourt-Brown S, Ramus F, van der Lely HJK. The link between prosody and language skills in children with specific language impairment (SLI) and/or dyslexia. International Journal of Language & Communication Disorders. 2009; 44:466–488. [PubMed: 19107654]

McCann J, Peppé S. Prosody in autism spectrum disorders: A critical review. International Journal of Language & Communication Disorders. 2003; 38:325–350. [PubMed: 14578051]

McCann J, Peppé S, Gibbon FE, O'Hare A, Rutherford MD. Prosody and its relationship to language in school-aged children with high-functioning autism. Journal of Language & Communication Disorders. 2007; 42:682–702.

Oller DK, Niyogi P, Gray S, Richards JA, Gilkerson J, Xu D, Yapanel U, Warren SF. Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. PNAS. 2010; 107:13354–13359. [PubMed: 20643944]

Paccia JM, Curcio F. Language processing and forms of immediate echolalia in autistic children. Journal of Speech and Hearing Research. 1982; 25:42–47. [PubMed: 7087424]

Paul R, Augustyn A, Klin A, Volkmar F. Perception and production of prosody by speakers with autism spectrum disorders. Journal of Autism & Developmental Disorders. 2005; 35:205–220. [PubMed: 15909407]

Paul R, Bianchi N, Augustyn A, Klin A, Volkmar FR. Production of syllable stress in speakers with autism spectrum disorders. Research in Autism Spectrum Disorders. 2008; 2:110–124. [PubMed: 19337577]

Peppé S. Why is prosody in speech-language pathology so difficult? International Journal of Speech-Language Pathology. 2009; 11:258–271.

Peppé S, McCann J, Gibbon F, O'Hare A, Rutherford M. Receptive and expressive prosodic ability in children with high-functioning autism. Journal of Speech, Language & Hearing Research. 2007; 50:1015–1028.

Peppé S, McCann J. Assessing intonation and prosody in children with atypical language development: The PEPS-C test and the revised version. Clinical Linguistics and Phonetics. 2003; 17:345–354. [PubMed: 12945610]

Peppé S, McCann J, Gibbon F, O'Hare A, Rutherford M. Assessing prosodic and pragmatic ability in children with high-functioning autism. Journal of Pragmatics. 2006; 38:1776–1791.

Rogers, SJ.; Williams, JHG. Imitation in autism: Findings and controversies. In: Rogers, SJ.; Williams, JHG., editors. Imitation and the social mind: Autism and typical development. New York: Guilford; 2006. p. 277-309.

Rutherford MD, Baron-Cohen S, Wheelwright S. Reading the mind in the voice: A study with normal adults and adults with Asperger syndrome and high functioning autism. Journal of Autism & Developmental Disorders. 2002; 32:189–194. [PubMed: 12108620]

Rutter, M.; Le Couteur, A.; Lord, C. Autism diagnostic interview-revised. Los Angeles: Western Psychological Services; 2003.

Schoen E, Paul R, Chawarska K. Phonology and vocal behavior in toddlers with autism spectrum disorders. Autism Research, online first. 2011

Semel, E.; Wiig, EH.; Secord, WA. Clinical evaluation of language fundamentals. 4th ed.. San Antonio: The Psychological Corporation; 2003.

Shriberg LD, Paul R, Black LM, van Santen JPH. The hypothesis of apraxia of speech in children with autism spectrum disorder. Journal of Autism and Developmental Disorders, online first. 2011

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 & Hearing Research. 2001; 44:1097–1115.

Stojanovik V, Setter J, Ewijk LV. Intonation abilities of children with williams syndrome: A preliminary investigation. Journal of Speech, Language & Hearing Research. 2007; 50:1606–1617.

Tager-Flusberg, H.; Paul, R.; Lord, C. Language and communication in autism. In: Volkmar, FR.; Paul, R.; Klin, A.; Cohen, DJ., editors. Handbook of autism and pervasive developmental disorders: Diagnosis, development, neurobiology, and behavior. Third ed.. Hoboken, New Jersey: John Wiley & Sons Inc; 2005. p. 335-364.

Van Santen JPH, Prud'hommeaux ET, Black LM, Mitchell M. Computational prosodic markers for autism. Autism. 2010; 14:215–236. [PubMed: 20591942]

Warren SF, Gilkerson J, Richards JA, Oller DK, Xu D, Yapanel U, Gray S. What automated vocal analysis reveals about the vocal production and language learning environment of young children with autism. Journal of Autism and Developmental Disorders. 2011; 40:555–569. [PubMed: 19936907]

Wechsler, D. Wechsler abbreviated scale of intelligence. San Antonio: Pearson; 1999.

Wells B, Peppé S. Intonation abilities of children with speech and language impairments. Journal of Speech, Language, and Hearing Research. 2003; 46:5–20.

Wells B, Peppé S. Intonation development from five to thirteen. Journal of Child Language. 2004; 31:749–778. [PubMed: 15658744]

Werner E, Dawson G, Osterling J. Brief report: Recognition of autism spectrum disorder before one year of age: A retrospective study based on home videotapes. Journal of Autism and Developmental Disorders. 2000; 30:157–162. [PubMed: 10832780]

Wetherby AM, Woods J, Allen L, Cleary J, Dickinson H, Lord C. Early indicators of autism spectrum disorders in the second year of life. Journal of Autism & Developmental Disorders. 2004; 34:473–493. [PubMed: 15628603]

Young EC, Diehl JJ, Morris D, Hyman SL, Bennetto L. The use of two language tests to identify pragmatic language problems in children with autism spectrum disorders. Language, Speech, and Hearing Services in Schools. 2005; 36:62–72.
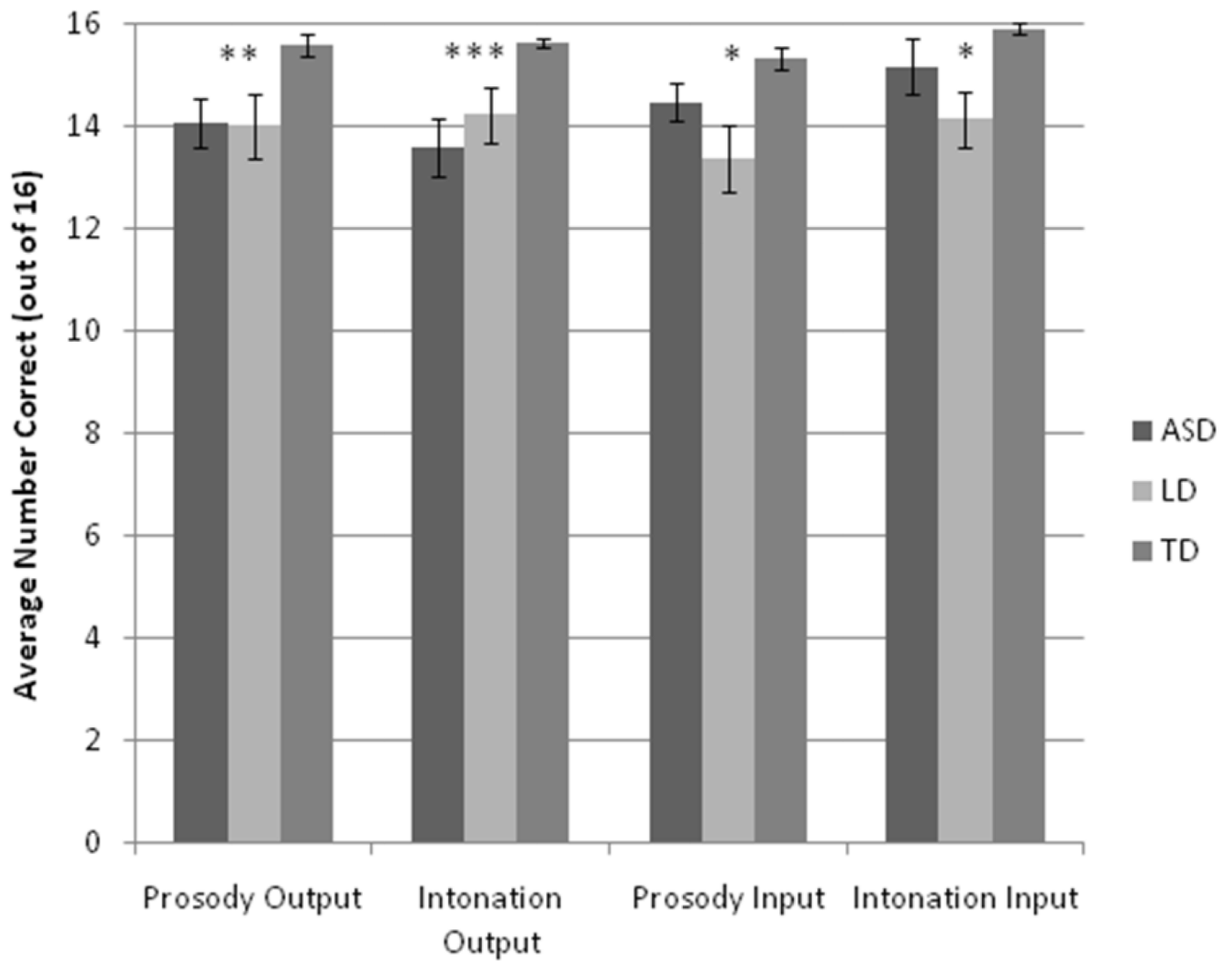
Figure 1. Number correct on behavioral responses from PEPS-C Form tasks
Note: Statistical tests represent Omnibus ANOVA for each subtest, or the Welch's test where statistically appropriate. Error bars represent standard error of the mean. PEPS-C = Profiling Elements of Prosodic Systems in Children (Peppé & McCann, 2003).
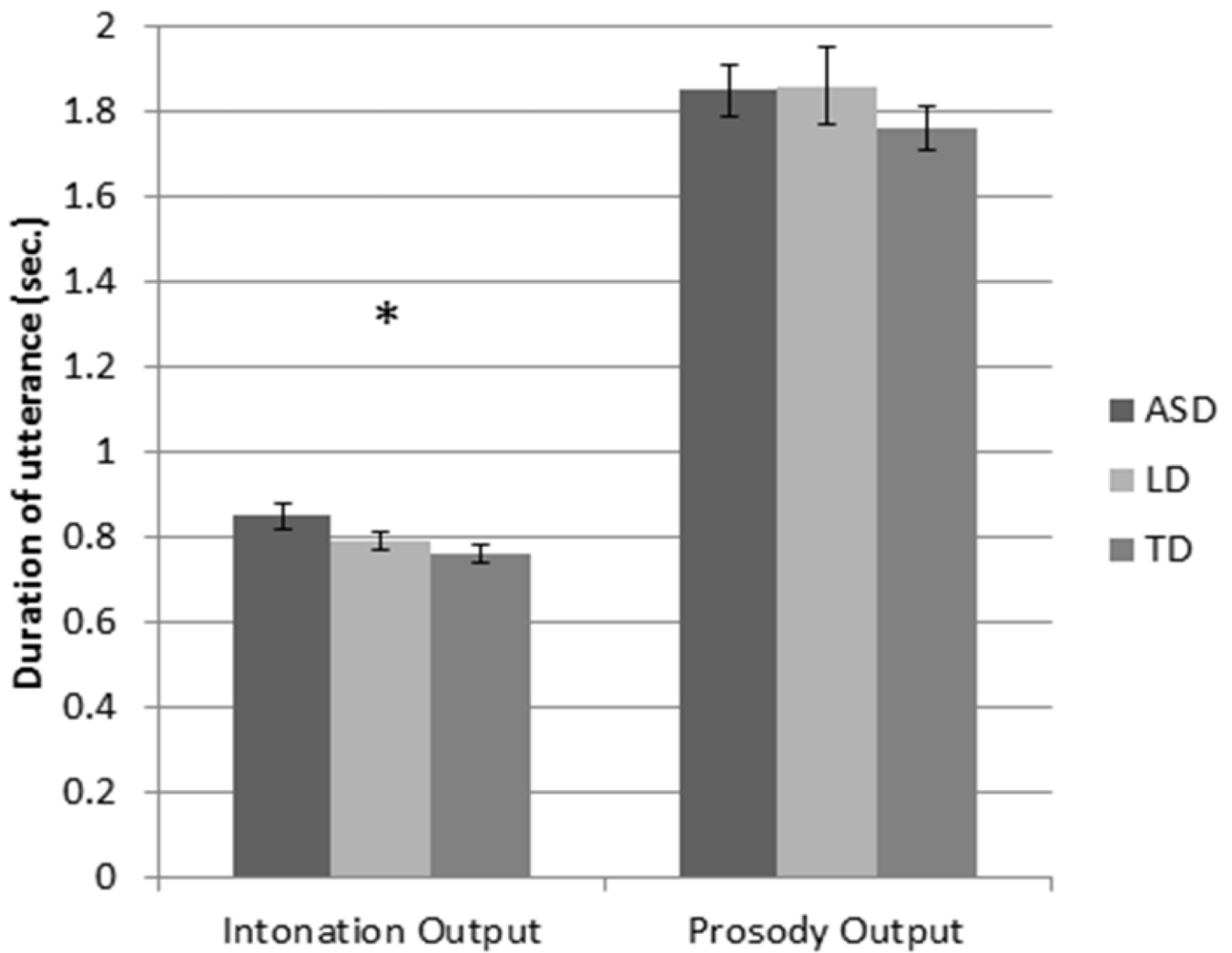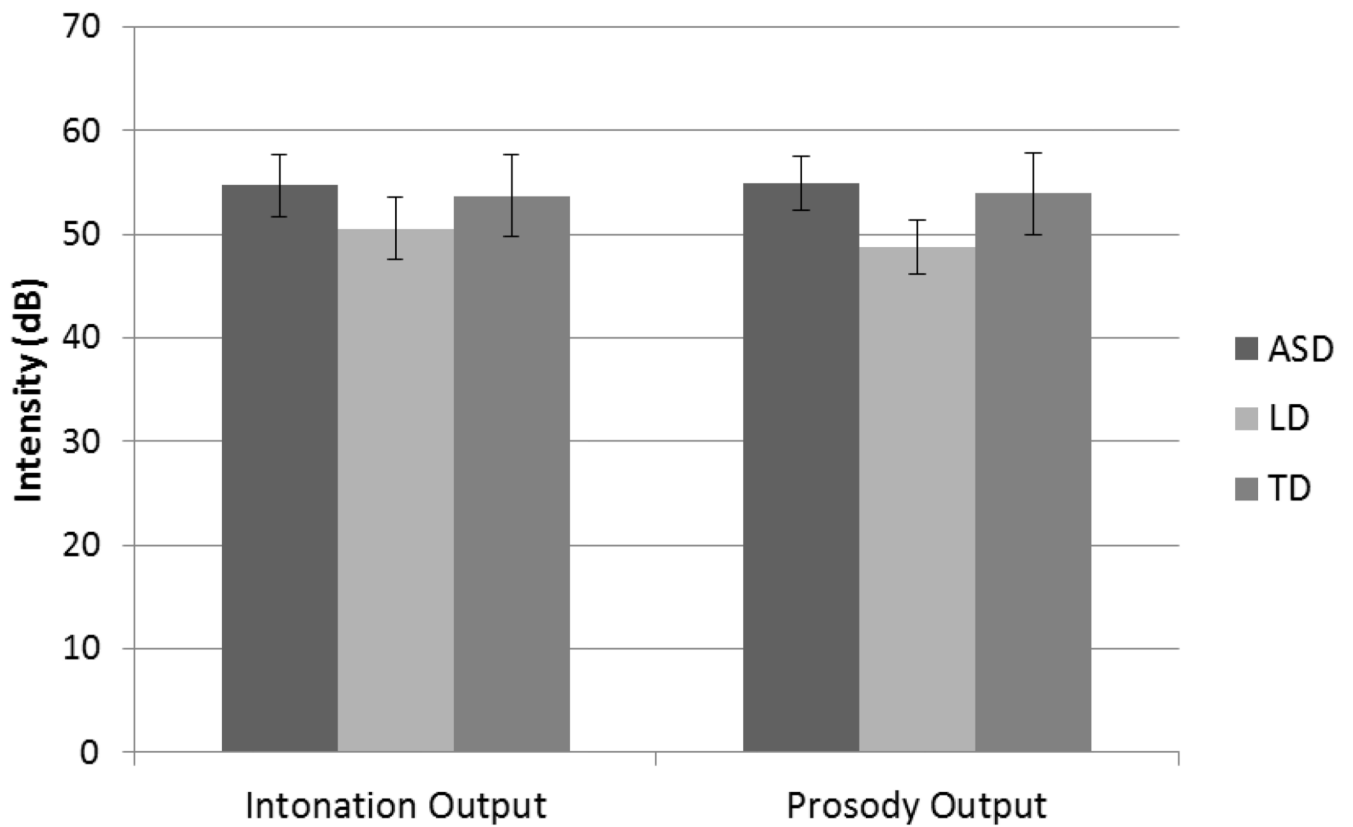* *p*<.05
** *p*<.01
*** *p*<.001

Figure 2. Average utterance duration, organized by task, for each participant group
Note: Statistical tests represent Omnibus ANOVA for each subtest, or the Welch's test
where statistically appropriate. Error bars represent standard error of the mean. PEPS-C =
Profiling Elements of Prosodic Systems in Children (Peppé & McCann, 2003).
* *p*<.05

Figure 3. Average utterance intensity, organized by task, for each participant group
Note: Statistical tests represent Omnibus ANOVA for each subtest, or the Welch's test where statistically appropriate. Error bars represent standard error of the mean. PEPS-C = Profiling Elements of Prosodic Systems in Children (Peppé & McCann, 2003).
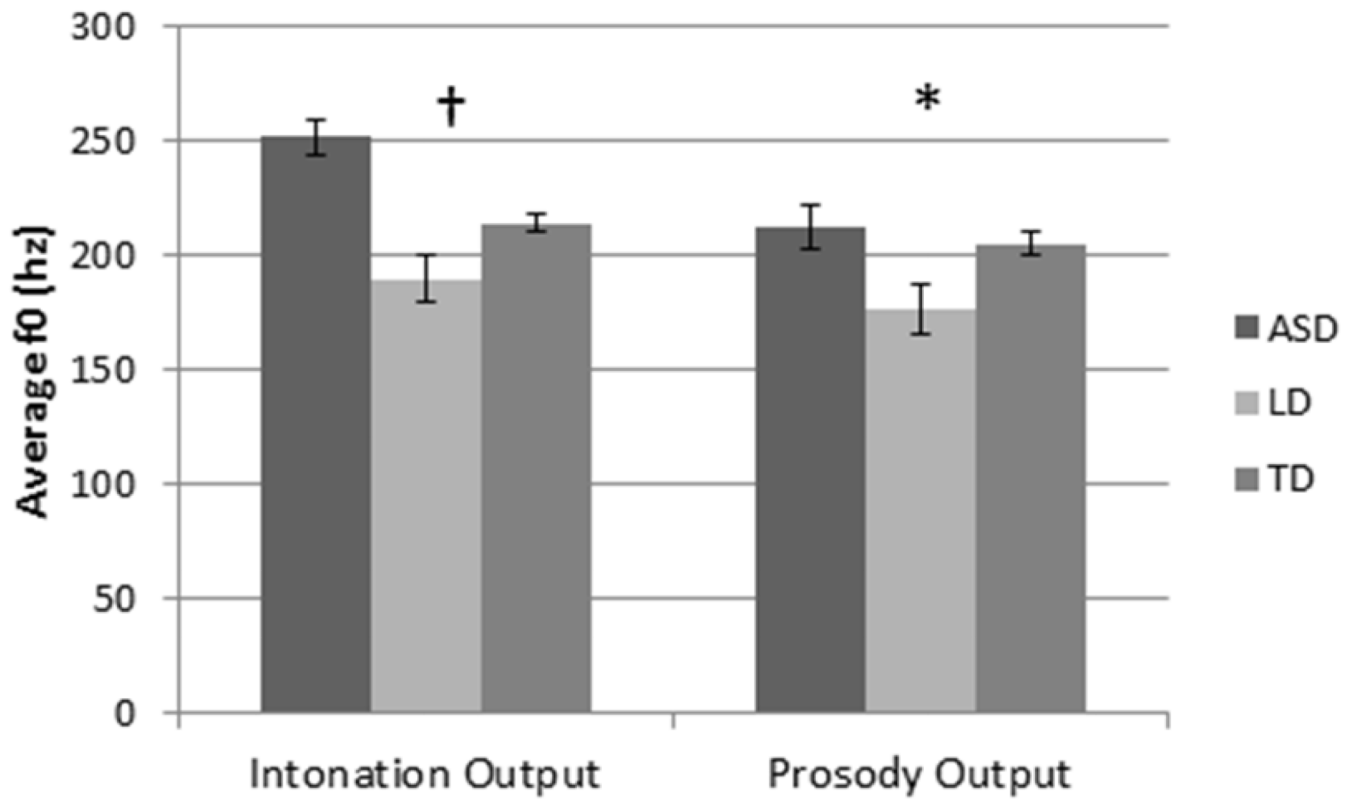
Figure 4. Average fundamental frequency, organized by task, for each participant group
Note: Statistical tests represent Omnibus ANOVA for each subtest, or the Welch's test where statistically appropriate. Error bars represent standard error of the mean. PEPS-C = Profiling Elements of Prosodic Systems in Children (Peppé & McCann, 2003).
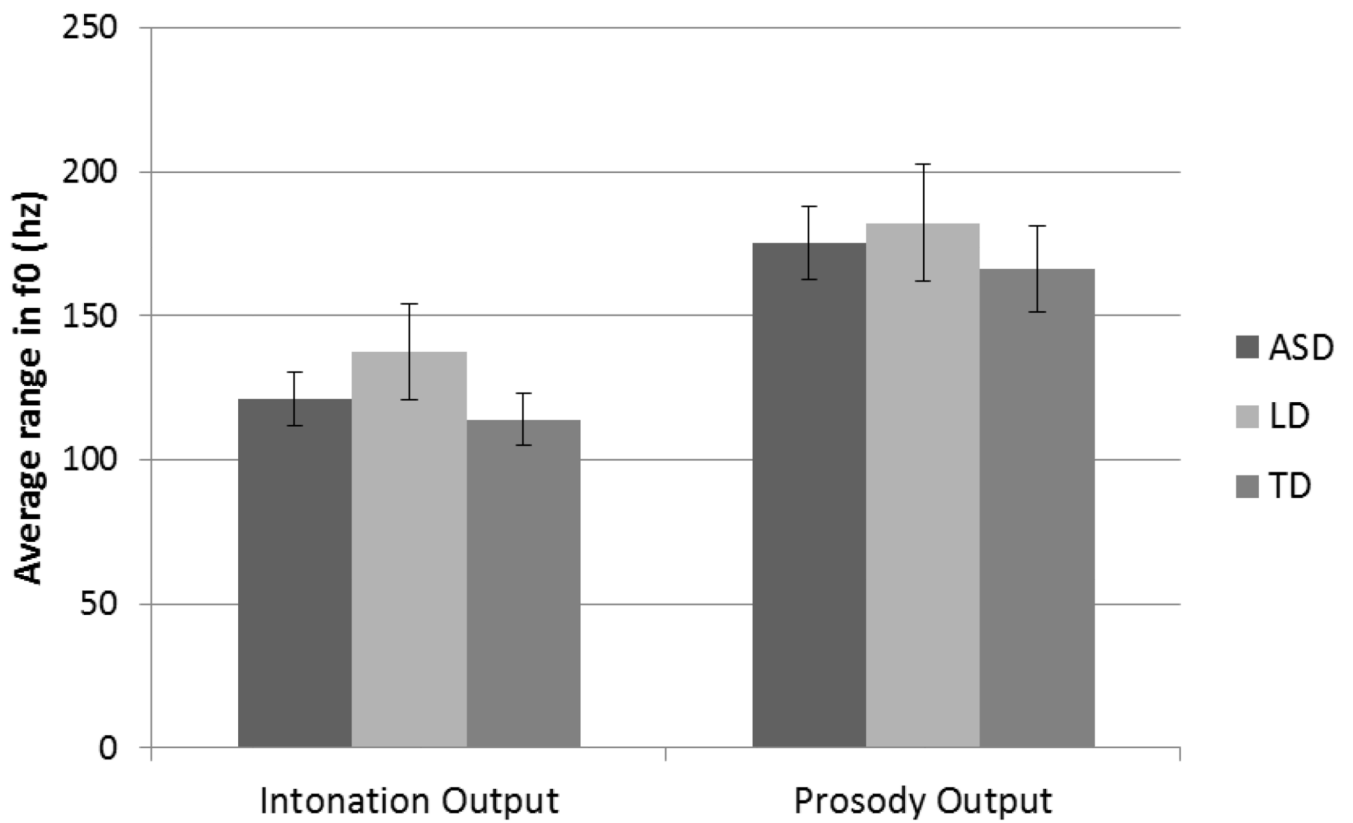* $p<.05$
† $p<.10$

Figure 5. Fundamental frequency range for correct responses, organized by task, for each participant group
Note: Statistical tests represent Omnibus ANOVA for each subtest, or the Welch's test where statistically appropriate. Error bars represent standard error of the mean. PEPS-C = Profiling Elements of Prosodic Systems in Children (Peppé & McCann, 2003).

Figure 6. Standard deviation of fundamental frequency for correct responses, organized by task, for each participant group
Note: Statistical tests represent Omnibus ANOVA for each subtest, or the Welch's test where statistically appropriate. Error bars represent standard error of the mean. PEPS-C = Profiling Elements of Prosodic Systems in Children (Peppé & McCann, 2003).
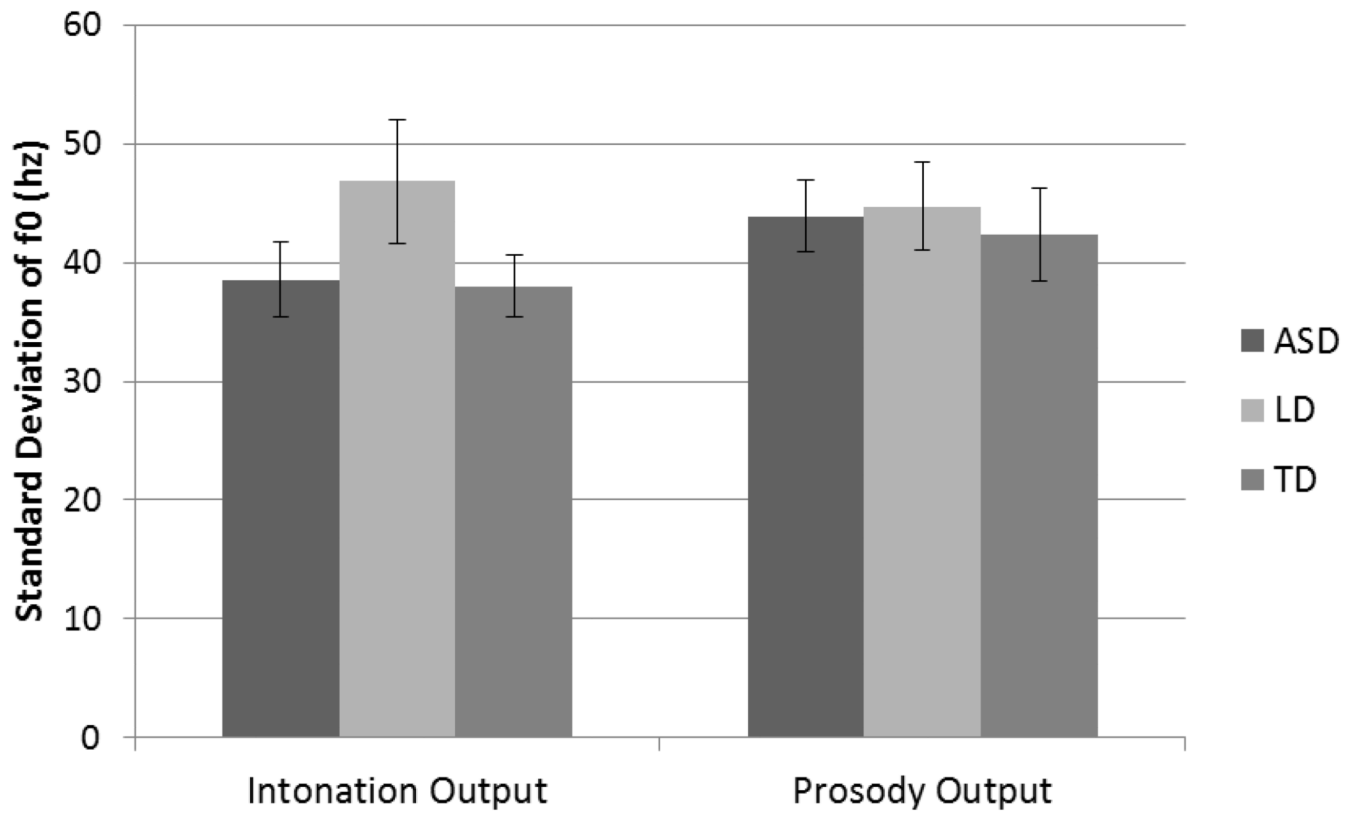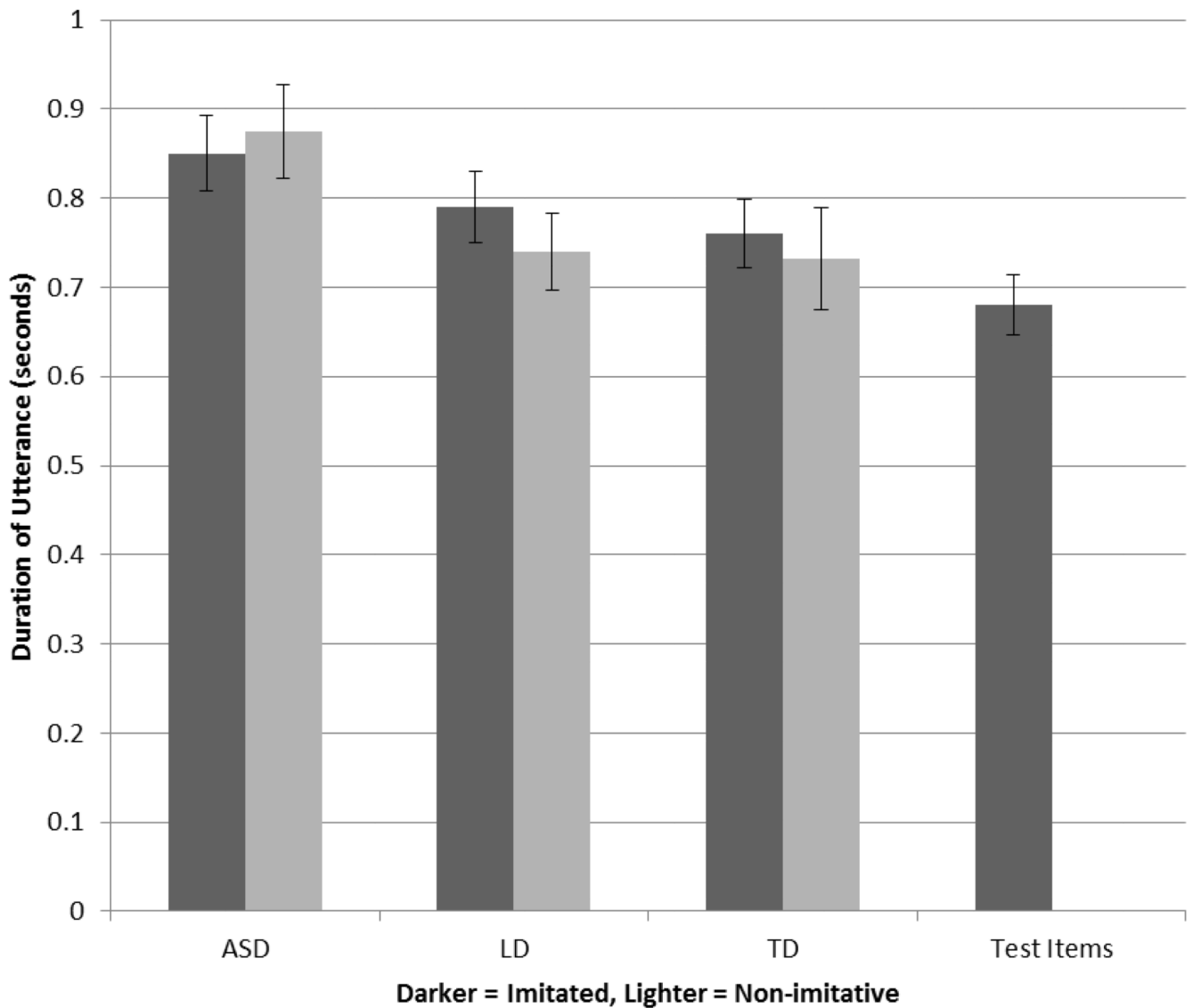
Figure 7. Comparison of imitated, non-imitative, and test item durations
Note: Error bars represent standard error of the mean. PEPS-C = Profiling Elements of Prosodic Systems in Children (Peppé & McCann, 2003). Data from non-imitative (spontaneous productions) taken from Diehl & Paul (in press).

Table 1

Descriptive characteristics of the sample

	ASD M (SD) [range]	LD M (SD) [range]	TD M (SD) [range]	F	p
N	24	16	22		
Gender (M:F)	16:8	12:4	15:7		
Chronological Age	12.31 (2.32) [8–16]	12.99 (2.25) [9–16]	12.21 (2.64) [9–17]	.55	.58
CELF-IV Receptive Language Index	93.67 (19.49) [58–121]	88.73 (17.63) [58–119]		.64	.43
CELF-IV Expressive Language Index	100.54 (16.22) [75–126]	90.00 (14.95) [65–114]		4.13	.05
CELF-IV Core Language	97.21 (18.61) [67–132]	88.94 (16.02) [60–117]		.82	.37
Nonverbal IQ[a]	103.61 (17.14) [75–133]	96.85 (11.13) [67–109]		1.54	.22

Note. CELF-IV=Clinical Evaluation of Language Fundamentals, 4th edition.

[a] Nonverbal IQ was measured using either the Performance IQ scaled score from the Wechsler Abbreviated Scale of Intelligence (Wechsler, 1999) or the Nonverbal IQ scale from the Differential Ability Scales (Elliot, 1990). Some nonverbal IQ scores were missing for participants, so for these analyses N=18 for the ASD group, and N=13 for the LD group.

Table 2

Description of PEPS-C Form Tasks

	Child Task	**Examiner Score**
Input		
"Intonation"	Participants heard a pair of words, but the speech was filtered such that all that the participants could hear was the intonation of the word. Participants determined whether the two patterns were the same or different. Patterns were similar to the ones used in Intonation Output and the PEPS-C Affect and Turn-End Type Function tasks.	Response scored automatically by the computer program.
"Prosody"	Participants heard a pair of short phrases, but the speech was filtered such that all that the participants could hear was the prosody of the phrase. Participants determined whether the two patterns were the same or different. Patterns were similar to the ones used in Prosody Output and on the PEPS-C Chunking Function tasks.	Response scored automatically by the computer program.
Output		
"Intonation"	Participants heard a single word, said with a specific intonation pattern, and had to imitate the word and intonation pattern. Patterns were similar to the ones used in Intonation Input and the PEPS-C Affect and Turn-End Type Function tasks.	Examiner judged whether imitated phrase was good (1 point), fair (.5 points) or poor (0 points).
"Prosody"	Participants heard short phrases, such as "pink and red and black socks," that were said with specific prosodic cues, (e.g., "pink - and red and black socks," or "pink and red - and black socks", where the "-" represents a subtle pause), and had to repeat the phrase. Patterns were similar to the ones used in Prosody Input and on the PEPS-C Chunking Function tasks.	Examiner judged whether imitated phrase was good (1 point), fair (.5 points) or poor (0 points).