JSLHR

Research Article

Automated Analysis of Child Phonetic Production Using Naturalistic Recordings

Dongxin Xu,[a,b] Jeffrey A. Richards,[a] and Jill Gilkerson[a,b]

Purpose: Conventional resource-intensive methods for child phonetic development studies are often impractical for sampling and analyzing child vocalizations in sufficient quantity. The purpose of this study was to provide new information on early language development by an automated analysis of child phonetic production using naturalistic recordings. The new approach was evaluated relative to conventional manual transcription methods. Its effectiveness was demonstrated by a case study with 106 children with typical development (TD) ages 8–48 months, 71 children with autism spectrum disorder (ASD) ages 16–48 months, and 49 children with language delay (LD) not related to ASD ages 10–44 months.
Method: A small digital recorder in the chest pocket of clothing captured full-day natural child vocalizations, which were automatically identified into consonant, vowel, nonspeech, and silence, producing the average count per utterance (ACPU) for consonant and vowel.
Results: Clear child utterances were identified with above 72% accuracy. Correlations between machine-estimated and human-transcribed ACPUs were above 0.82. Children with TD produced significantly more consonants and vowels per utterance than did other children. Children with LD produced significantly more consonants but not vowels than did children with ASD.
Conclusion: The authors provide new information on typical and atypical language development in children with TD, ASD, and LD using an automated computational approach.

Key Words: vocal development, phonetic development, typical development, language delay, autism, speech recognition, adult acoustic phonetic model

Research shows that different developmental issues affect the trajectory of a child's language development (Grizzle & Simms, 2005; Tager-Flusberg et al., 2009; Thomas, 2010) and that phonetic development is a major foundation of overall language development in young children (Vihman, 1996). A detailed analysis of early vocalizations may provide useful diagnostic information not only about language development per se, but also about a child's overall developmental course.

Child phonetic development includes both phonetic perception and phonetic production. Although research examining early child speech and phonetic perception is important for a complete understanding of child vocalization, speech production, and language development (Kuhl, 2004; Kuhl & Meltzoff, 1996), we focused only on child phonetic production and used a novel automated approach to provide new information on early language development.

The basis of phonetic and phonological analysis is the categorization of child vocalizations, which includes both phonetic production and prelinguistic vocalization (Davis, MacNeilage, & Matyear, 2002; Dodd & McIntosh, 2010; Ferguson & Farwell, 1975; Kuhl & Meltzoff, 1996; McCune & Vihman, 2001; Nathani, Ertmer, & Stark, 2006; Oller, 1980, 2000; Pierrehumbert, 2003; Rescorla & Bernstein-Ratner, 1996; Serkhane, Schwartz, Boë, Davis, & Matyear, 2007; Stark, 1980; Stoel-Gammon, 1985; Vihman, 1996; Williams & Elbert, 2003). Child vocalizations have been categorized according to various characteristics—for example, by depicting them as *speech-like* or *nonspeech-like* (Kuhl, 2004; Kuhl & Meltzoff, 1996; Nathani et al., 2006; Oller, Eilers, Steffens, Lynch, & Urbano, 1994). Speech-like vocalizations can be further categorized into phonetic units, such as consonants, vowels, and/or syllables, as well as higher level words, phrases, and utterances (Dyson, 1988; Ingram, 2002; Ingram & Ingram, 2001; McCune & Vihman, 2001; Pierrehumbert, 2003; Rescorla & Bernstein-Ratner, 1996;

[a]LENA Research Foundation, Boulder, CO
[b]University of Colorado, Boulder

Correspondence to Dongxin Xu: dongxinxu@lenafoundation.org

Editor: Jody Kreiman
Associate Editor: Megha Sundara

Received February 12, 2013
Revision received November 26, 2013
Accepted May 5, 2014
DOI: 10.1044/2014_JSLHR-S-13-0037

Disclosure: The study and related research and development were funded by the LENA Research Foundation. All three authors, Dongxin Xu, Jeffrey A. Richards, and Jill Gilkerson, are employees of the LENA Research Foundation.

Stoel-Gammon, 1991).[1] Prelinguistic vocalization is relatively less well formed compared with phonetic production, tends to have higher variation, and generally can be sorted into vegetative sounds, fixed signals (such as cries), and protophones (Oller, 1980, 2000). Other categorization schemes have also been developed (Nathani et al., 2006; Stark, 1980).

Conventional phonetic and phonological analyses of child language development are based on human identification of consonants and vowels in a child's spoken words and have been studied extensively. Various measures are used to characterize vocal development, including consonant inventories, frequency of occurrence of consonants and vowels, phonological mean length of utterance, and so on (Davis et al., 2002; Ingram, 2002; Ingram & Ingram, 2001; McCune & Vihman, 2001; Rescorla & Bernstein-Ratner, 1996; Saaristo-Helin, Kunnari, & Savinainem-Makkonen, 2011; Saaristo-Helin, Savinainem-Makkonen, & Kunnari, 2006; Serkhane et al., 2007; Stoel-Gammon, 1985, 1991, 2010; Vihman, Keren-Portnoy, Bidgood, McGillion, & Whitaker, 2013; Williams & Elbert, 2003).

A detailed analysis of child phonetic and phonological development usually exploits data from both elicited and spontaneous samples. Elicited single words collected via picture-naming procedures, for example, can virtually exhaust all possible phonemes and phoneme combinations; however, some child vocalization phenomena, such as intelligibility and prosody, are prominent only in spontaneous or conversational speech (Saaristo-Helin, 2011; Stoel-Gammon, 2010). Spontaneous samples can be collected in laboratory, clinical, or natural home environment settings. Sometimes, naturalistic samples are more desirable (Tager-Flusberg et al., 2009.)

Studies of child development from phonetic and phonological perspectives have been widely conducted not only in children with typical development (TD) but also in children with specific language impairments, children with language delays, children with hearing impairments, children with autism, and children with other developmental disorders (Ertmer, 2001; Ingram, 2002; Oller, Eilers, Bull, & Carney, 1985; Stark, 1983; Stoel-Gammon, 1988; Tager-Flusberg et al., 2009; Vihman, 1996). Cross-group or cross-disorder comparisons are potentially informative about the different developmental processes for different groups or disorders (Thomas, 2010.)

Well-established methodologies have been developed for use in the phonetic and phonological analyses of the vocalizations of young children. Yet, despite the rigor with which such studies may be conducted, three key limitations have been identified: (a) the resource burden (practical, temporal, financial, etc.) of obtaining sufficiently large representative vocalization samples from young children generally limits studies to relatively brief sampling regimens; (b) the level of environmental control necessary to obtain recorded data of sufficient quality often dictates that studies be conducted in laboratory and clinical settings rather than under more naturalistic scenarios; and (c) the need for highly specialized training and expertise to collect and manually code or transcribe vocalizations reduces the number and extent of studies.

The above-mentioned limitations are interrelated. Incorporating naturalistic vocalization data usually requires much more frequent sampling and larger data sets because of the high degree of variation not only in the productions of interest, but also in confounding factors, such as noise and behavior in uncontrolled environments. Such factors traditionally create a bottleneck for phonetic development studies in naturalistic settings because of the need for laborious manual transcription and coding. Labor-intensive tasks substantially reduce the potential clinical applicability of such research, as the high cost and time-consuming requirements of sample collection and processing are too impractical to implement outside of a research setting.

With the advance of modern technologies in microelectronics, speech signal processing, and pattern recognition, new possibilities have emerged that can address the limitations of conventional approaches and provide efficient means to expand previous capabilities. In this study, we equipped child subjects with a small lightweight digital recorder for the purpose of recording their full-day natural vocal output and environment. Computational algorithms were used to identify child sound segments (traditionally termed "child utterances") in the full-day recordings. Phonetic units, such as consonants and vowels, in child sound segments were identified by speech recognition software. Average count per utterance (ACPU) for consonant and vowel was used to evaluate child vocal development. ACPU measures, which reflect general language development to some extent, are associated with the number of words spoken per utterance or the length or complexity of the words used.

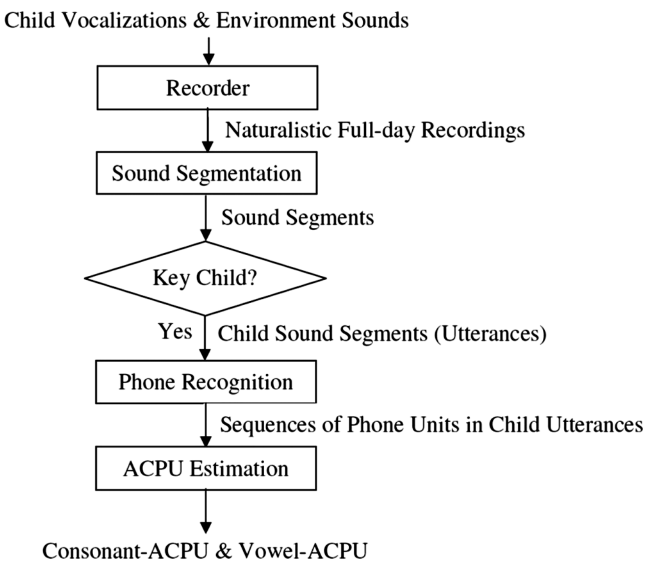
The current study included two parts. The first part evaluated the new automated computational approach by comparing its results with those observed using conventional manual approaches. The second part explored the effectiveness of the new approach using a case study of the developmental trends and child group differences among TD children and those with autism spectrum disorder (ASD) and language delay (LD) not related to ASD. The following sections of the article are organized accordingly.

Part 1: The Automated Approach and Its Validation

A diagram of the new approach is shown in Figure 1. It consists of three major stages: (a) naturalistic recording, (b) automated sound segmentation to identify child sound segments (utterances) in audio streams of recordings, and

[1]Due to the developing and uncertain nature of young child vocalization, terms such as *consonant-like, vowel-like, speech-like, nonspeech-like, silence-like,* etc., are often used to represent "fuzzy" child vocal categories. Another reason to use such terms is that the categorization used in this study is based on automated algorithmic speech recognition and the resulting macrostatistics of large data samples. For simplicity, we refer to these categories as *consonant*, *vowel*, *speech, nonspeech,* and *silence.*

Figure 1. Diagram showing the automated approach for sample collection and data processing. ACPU = average count per utterance.



(c) recognition of phonetic units in child utterances and estimation of Consonant-ACPU and Vowel-ACPU.

Naturalistic Recording

Naturalistic full-day audio recordings were obtained using the LENA digital language processor (DLP), a small lightweight digital recorder. In the current study, DLPs were worn by children in the chest pocket of specially designed clothing to record vocalizations and surrounding sounds in the natural environment. The DLP can record unobtrusively and continuously for up to 16 hr. Additional technical specifications about the DLP can be found in Xu et al. (2008) and Oller et al. (2010).

Automated Sound Segmentation

The first step in the segmentation process involved locating and identifying different types of sound segments in daylong audio streams. Eight types of sound segments, which were based on previously trained statistical models, were delineated in the recordings: *key child* (i.e., the child who wears the DLP), *other child, adult male, adult female, overlapped sounds*, *noise, electronic media* (such as TV), and *silence.*[2] The seven nonsilence types of sound segments were further classified as "clear" or "not clear" via likelihood ratio tests between the winning model and the silence model, which resulted in a total of 15 categories.

The algorithmic processing techniques used in the current approach are presented conceptually here and described in more detail in Xu et al. (2008) and Xu, Gilkerson, Richards, Yapanel, and Gray (2009). The basic segmentation algorithm can be conceptualized as a "searching" process to identify the optimal result. With statistical models of sound categories, the likelihood of any possible segmentation result (or segment sequence) can be calculated. The conceptual brute force algorithm can list all possible segment sequences and calculate their likelihood values, and then select the one with maximum likelihood. The actual algorithm simply uses a more efficient method (i.e., dynamic programming) to find the segment sequence with maximum likelihood. It does so by eliminating improbable candidates at every stage in the searching process, thereby resulting in far less computation and the guarantee of finding the optimal solution. This algorithm was implemented in the LENA software, which takes audio recordings and child age/gender information as input and produces sound segment sequences with time stamps (Xu et al., 2008).

The validation of the sound segmentation algorithm was based on two test data sets: Test-Set-70 and Test-Set-94 (see Table 1). A confusion matrix, which shows the percentages of agreement between the machine and the human transcriber, is presented for each data set.

Test-Set-70 included 70 children ages 2–36 months; each month-age had two children, and each child had 1 hr of recorded data (i.e., a total of 70 hr of recorded data). In Table 1, the first row shows that 76.1% of data transcribed by human listeners as "clear key child sound segments" was correctly identified by the machine and that 23.9% of data was identified as "other sound segments." Clear key child sound segments are considered approximations to clear key child utterances. More detailed confusion matrices and information can be found in Xu et al. (2008).

To replicate the above findings, we recently added 24 additional children between the ages of 37 and 48 months (2 children per month-age and 0.5 hr of recording data per child) to the Test-Set-70 data set. The expanded data set (i.e., Test-Set-94) consisted of 94 children ages 2–48 months. The confusion matrix for Test-Set-94 showed an accuracy rate of more than 72% for clear key child sound segments, which is similar to that for Test-Set-70.

Recognition of Phonetic Units in Child Utterances and Estimation of ACPUs

Once different sound segments in audio streams have been identified, they can be further processed for various purposes. In our study, clear key child segments (utterances) were further processed for phonetic analysis. Open source Sphinx speech recognition software, which can be downloaded from the Sphinx website (http://cmusphinx.sourceforge.net/), was used to recognize phonetic units in child utterances. Hub4 adult acoustic phonetic models were used in the Sphinx speech recognition algorithms, which were trained

[2]Naturalistic recordings may include silent periods (*silence*), which differ from within-utterance silence. Within-utterance silence, which has not been addressed in the current study, is usually short in duration and may be associated with pauses, phrase boundaries, consonant production, prosody, and so on.

Table 1. Performance of sound segmentation: Confusion matrixes.

Human transcribed	Machine identified: Clear key child segments	Machine identified: Other sound segments
Test-Set-70 (70 hr, 70 children, ages 2–36 months)		
Clear key child segments	76.1%	23.9%
Other sound segments	3.9%	96.1%
Test-Set-94 (82 hr, 94 children, ages 2–48 months)		
Clear key child segments	72.5%	27.5%
Other sound segments	4.0%	96.0%

from a corpus of well-articulated broadcast news speech (Seymore et al., 1998; Sphinx website). Thus, the acoustic models used were based on well-pronounced adult speech. This process is similar to "relational analysis" of conventional approaches (Saaristo-Helin et al., 2011; Stoel-Gammon, 2010), in which adult models or targets are compared with child samples. The difference is that the comparison here is at a phonetic level, not a word level, and is based on acoustic similarity calculated by machine algorithms, not human judgment. The goal of phone recognition was to characterize (a) 39 English phonetic units (24 consonants and 15 vowels), (b) within-utterance silence, and (c) six additional categories of other nonspeech sound elements, such as hesitation, coughing, noise, lip smacking, etc. For each utterance processed, the software provided phone labels (for the 46 categories, = 39 + 1 + 6) and time marks, which, in turn, served as the basis for further analysis. Sphinx phone categories can be further merged into four types: consonant, vowel, nonspeech, and silence. In the current study, Consonant-ACPU and Vowel-ACPU were estimated for each recording.

The interface and algorithm parameters of Sphinx software may need adjustment to accommodate specific cases. The phone recognition with Sphinx software was integrated into the LENA software to ensure that data processing was fully automated without any human involvement.

Consonants and vowels identified by machine in this way cannot be the same as those identified by a human; however, high correlations were expected between the two methods. The validation of the recognition of phonetic units in the current study was directly conducted on the final ACPU measures rather than the details of phone recognition.

A small data set was phonetically transcribed prior to conducting the current study for other purposes. Human transcribers were asked to transcribe phonetic units based on words perceived with context information from audio streams of 62 recordings. The transcribers selected child utterances if all sound units of the utterances could be interpreted as either consonants or vowels without nonspeech units. The time marks of phonetic units were not required; thus, the transcription contained only sequences of labels of consonants and vowels without time marks. An average of 38.3 child utterances were transcribed for each recording. A histogram showing the number of transcribed child utterances in each recording is presented in Figure 2.

The above-mentioned data set was used to validate the automated ACPU measures. Human-transcribed ACPUs included only Consonant-ACPU and Vowel ACPU. Machine estimates of ACPUs were calculated based on the phonetic labels produced by machine and included not only Consonant-ACPU and Vowel-ACPU, but also Nonspeech-ACPU. With Nonspeech-ACPU, we can check its impact on machine-based Consonant-ACPU and Vowel-ACPU.

Figure 3 shows scatter plots of Consonant-ACPU and Vowel-ACPU (machine vs. human-transcribed estimates). The green lines represent the points at which machine estimates were equal to human-transcribed estimates, which helps to illustrate the difference between machine and human measures. The scatter plots show that both the machine Consonant-ACPU and Vowel-ACPU were highly correlated with their corresponding human-transcribed estimates; however, both were underestimated. Table 2 and Table 3 show that the correlation between the machine and the human Consonant-ACPUs was 0.85 ($p < .001$) and that for the Vowel-ACPU was 0.82 ($p < .001$); however, t tests indicate that the difference between the human and the machine measures was significantly greater than 0, with a mean of 0.56 and 0.57 for consonant and vowel, respectively.

To understand the lower machine counts for both consonants and vowels and the impact of machine recognized nonspeech, we further compared variables between

Figure 2. Number of transcribed child utterances in each recording of the test set for the validation of ACPU measures.

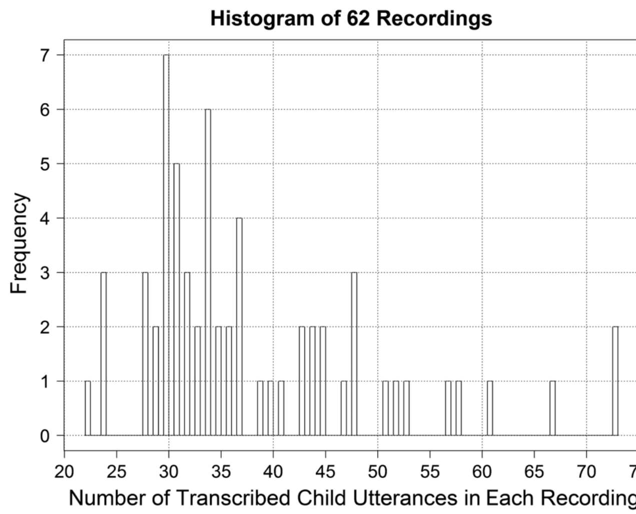
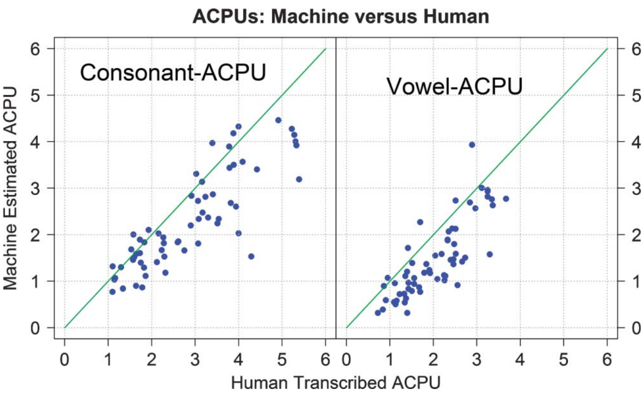


Figure 3. Scatter plots of machine-estimated ACPUs versus human-transcribed ACPUs. A blue dot represents one of 62 recordings. Green lines represent $y = x$ diagonal lines, which help to show that machine-estimated ACPUs are highly correlated with human transcribed ACPUs but are underestimated comparatively.



machine and human in terms of total counts or sums of ACPUs. Specifically, on the human side, there was only one variable: the sum of Consonant-ACPU and Vowel-ACPU. On the machine side, there were two variables: the sum of Consonant-ACPU and Vowel-ACPU and the sum with additional Nonspeech-ACPU. Figure 4 shows the two corresponding scatter plots, and Table 2 and Table 3 show the corresponding analyses. Both machine variables were again significantly and highly correlated with the human variable (0.86 and 0.89, respectively). As expected, if machine-recognized nonspeech units were excluded, the sum of machine ACPUs was underestimated relative to the human variable; however, if nonspeech units were included, the sum of the three machine ACPUs became significantly larger than the human variable.

Part 2: An Application Example Using the New Automated Approach

One purpose of the current study was to test the effectiveness of the proposed automated approach by examining the development trends and group differences of three diagnostic child groups. In the following section, we begin with a discussion of our prestudy predictions and then describe the methods and results of the case study we conducted.

Prediction

The phonetic development of children with ASD has been compared with that of TD children as well as with children at high and low risk of ASD, using conventional approaches (Paul, Fuerst, Ramsay, Chawarska, & Klin, 2011; Tager-Flusberg et al., 2009). In general, within an age range, children with ASD or those at high risk of ASD produced fewer consonants and other phonetic units, such as canonical syllables, than did their TD counterparts or those in groups at low risk of ASD. Paul et al. (2011) showed that children in both low- and high-risk groups produced more phonetic units as their age increased. Based on results from conventional approaches, we predicted and hypothesized that both measures of Consonant-ACPU and Vowel-ACPU would show significant differences among child groups under study and demonstrate significant developmental trends based on child age.

Participants

The participants and audio recordings included in this study were drawn from a previously reported sample, which is briefly summarized here; additional demographic data and other details can be found in Oller et al. (2010). Participants included 106 children with TD ages 8–48 months, 71 children diagnosed with ASD ages 16–48 months, and 49 children diagnosed with LD (not related to ASD) ages 10–44 months. In contrast to the previously reported study, recordings of children with ASD were excluded on days when they attended therapy sessions (as confirmed by the parents) for fair comparison of child behavior in the naturalistic environment. In total, 1,363 recordings were included, representing over 1,000 hr of child vocalizations and nearly 3.5 million individual child sound segments. Table 4 provides a summary by diagnostic group, and Figure 5 shows the distribution of recordings across different age-months for different children. We appreciate that longitudinal data are lacking in the ASD group and should be considered in subsequent analyses. Overall, the number and lengths of recordings in this data set demonstrate the efficiency of the naturalistic data collection using an automated approach.

Table 2. Performance of machine-estimated ACPU measures: Correlations between machine-estimated ACPUs and human-transcribed ACPUs.[a]

Human transcribed	Machine estimated	Correlation	95% confidence interval	*t*
Consonant-ACPU	Consonant-ACPU	0.85*	[0.76, 0.91]	12.6
Vowel-ACPU	Vowel-ACPU	0.82*	[0.72, 0.89]	11.2
Consonant-ACPU + Vowel-ACPU	Consonant-ACPU + Vowel-ACPU	0.86*	[0.78, 0.92]	13.2
Consonant-ACPU + Vowel-ACPU	Consonant-ACPU + Vowel-ACPU + Nonspeech-ACPU	0.89*	[0.83, 0.93]	15.2

Note. ACPU = average count per utterance.

[a]$N = 62$.

*$p < .001$.

Table 3. Performance of machine-estimated ACPU measures: Differences between human-transcribed variables and machine estimates.[a]

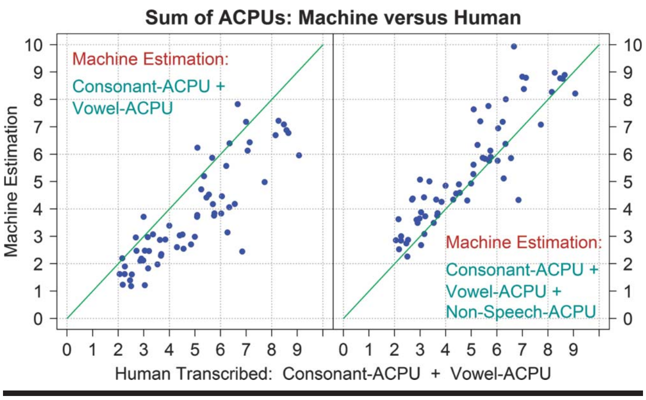
Human transcribed	Machine estimated	Correlation	95% confidence interval	t
Consonant-ACPU	Consonant-ACPU	0.56*	[0.40, 0.72]	6.9
Vowel-ACPU	Vowel-ACPU	0.57*	[0.44, 0.69]	9.3
Consonant-ACPU + Vowel-ACPU	Consonant-ACPU + Vowel-ACPU	1.13*	[0.87, 1.38]	8.9
Consonant-ACPU + Vowel-ACPU	Consonant-ACPU + Vowel-ACPU + Nonspeech-ACPU	–0.54*	[–0.78, –0.31]	–4.6

[a]One-sample t test for difference variable (Human_Variable – Machine_Variable), $N = 62$.
*$p < .001$.

Descriptive Statistics and Data Visualization

The new approach automatically estimated Consonant-ACPU and Vowel-ACPU for each recording. The group means and standard errors were estimated for each child group and each child month-age. These group means and standard errors were used as descriptive statistics for data visualization purpose only. The statistical analyses were based on the original ACPU measures of each recording. As shown in Figure 5, data were scarce for some month-ages, and some degree of smoothing was done to partially ameliorate this problem. We first chose the month-age ranges based on data availability for each child group as follows: (a) TD from 13 to 47 months; (b) LD from 15 to 38 months; and (c) ASD from 24 to 47 months. Second, to smooth results, we used a ±5-month moving inclusion range (or window) for each age. For example, the 30-month value represented the 25- to 35-month range, and in the TD group, the center month of each window moved from 13 to 47 months. Similar moving month-windows were applied to the LD and ASD groups. Within each month-window, child measures (scores) were obtained by directly averaging the raw recording measures for each child. Child group means and standard errors for each month were then estimated using child measures within the month-window. Group means and standard errors are plotted in Figure 6. The plots demonstrate the increasing trends of both Consonant-ACPU and Vowel-ACPU, especially in the TD group. The following group differences are also shown in the plots: (a) the differences between the TD group and the other two groups for both Consonant- and Vowel-ACPUs increase as children age, (b) the group differences are clearer for Consonant-ACPU than Vowel-ACPU, and (c) Vowel-ACPUs for the LD and ASD groups are basically overlapping, without much group difference shown in the plot.

Figure 4. Scatter plots of the sum of machine-estimated ACPUs versus the sum of human-transcribed ACPUs. A blue dot represents one of 62 recordings. Green lines represent $y = x$ diagonal lines, which help to show high correlations between machine and human variables and the differences between them. The scatter plots indicate that the total machine count for consonants and vowels is underestimated relative to the total human count of consonants and vowels. When including machine-recognized nonspeech units, the total machine count exceeds the human count.



Statistical Analyses

Statistical analyses were conducted to test the hypotheses pertaining to both the developmental trends and group differences of ACPU measures. We used correlational analyses for the developmental trends with both theoretical estimation and bootstrap simulation. Analysis of variance (ANOVA) and t tests were used for the analyses of child group differences with age-standardized z scores of ACPU measures. Additionally, we included a longitudinal analysis with the hierarchical linear model (HLM) (Singer & Willett, 2003) for this data set (see the Appendix). From the HLM analysis, we expected some additional information but essentially consistent results. In the following discussion, we focus on the age correlation analysis and the group differences analyses. All analyses, including the descriptive statistics and data visualization mentioned above, used the R statistical programming language (R Development Core Team, 2011).

Correlations With Child Age

Child developmental trends can be examined by correlating ACPU measures with child chronological month-age. One challenge we faced was treating and weighting children equally despite differences in the number of recordings. To deal with this issue, we used a weighted correlation to combine a child's multiple recordings in the analysis so that each child was weighted equally and thus treated equally; all of the raw recording ACPU measures were directly involved in the analysis. Specifically, if a child had n recordings, each recording was assigned a weight of $1/n$, resulting in an overall weight of 1. Weighted correlations were calculated in a way that was similar to a standard correlation

Table 4. Participant demographics.

Variable	TD sample	LD sample	ASD sample	Total
Participants	106	49	71	226
Age range (months)	8–48	10–44	16–48	8–48
Recordings	802	333	228	1,363
Child hours	634	224	159	1,017
Child utterances[a] (segments) in millions	2.15	0.75	0.53	3.43
Phonetic analysis units in millions[b]	8.42	2.65	1.82	12.89

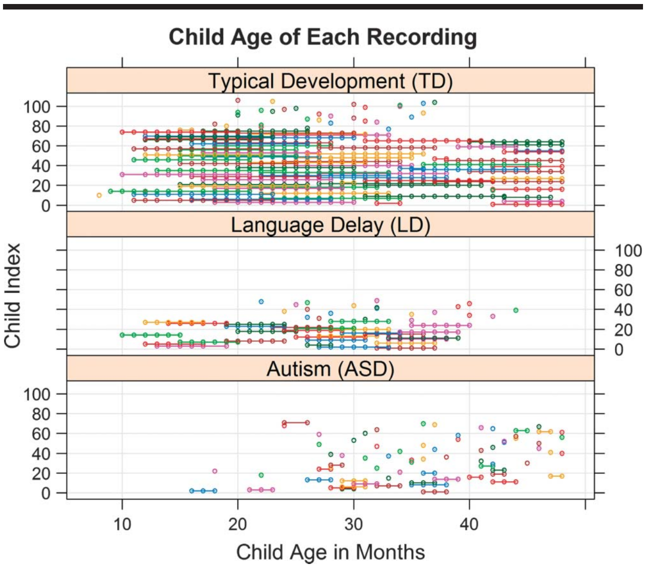
Note. TD = typical development; LD = language delay; ASD = autism spectrum disorder.

[a]Child sound segments in naturalistic recordings were identified by machine algorithm using the new automated method and were considered an approximation to child utterances. [b]Phonetic analysis units included machine-recognized consonants, vowels, nonspeech sound units, and within-utterance silences. (The focus of the current study was on consonants and vowels.)

by using ACPU measures of recordings, child month-ages, and weights. Statistical significances were determined by both theoretical estimation (Cohen, Cohen, West, & Aiken, 2003) and bootstrap simulation (Chernick & LaBudde, 2011), methods that are complementary to some extent; higher *p* values were chosen to ensure the rigor of the analyses. Bootstrap simulation was also used to test the significance of the group differences of the weighted correlations.

Table 5 and Table 6 show the results of the weighted correlation analysis between ACPUs and child chronological age. The TD group demonstrated significant and relatively strong correlations (0.63 for Consonant-ACPU and 0.58 for Vowel-ACPU). Both the LD and ASD groups showed weak or no significant correlations (below 0.33) for both Consonant-ACPU and Vowel-ACPU. The relative weakness of the correlations in the LD and ASD groups was significant, according to the test on the group differences of the correlations, which means that developmental trends are slower in children in the LD and ASD groups than those in the TD group.

Figure 5. Distribution of the recordings across child age. The vertical axis represents child index (i.e., the within-group participant index). For example, 106 typically developing children were indexed from 1 to 106. A point represents one recording. A flat line connecting multiple points represents one child with multiple recordings. A single point without connecting lines represents one child with only one recording or one child with multiple recordings within 1 month. Color is used solely for the purpose of better representation. One child is associated with one color, but a color may be repeated for different children.



Group Difference Analysis With Age-Standardized z Scores

As shown in Figure 6, ACPU measures change significantly with child month-age, and a child may have multiple recordings across different months. To generate more interpretable overall measures that incorporate all of a child's multiple recordings and use such child measures for group difference analysis, it is first necessary to age-standardize the measures. For this purpose, we computed age-standardized

Figure 6. Developmental trends using descriptive statistics for means and standard errors of ACPU measures (machine estimates). Group differences can also be identified through data visualization, especially with Consonant-ACPU.

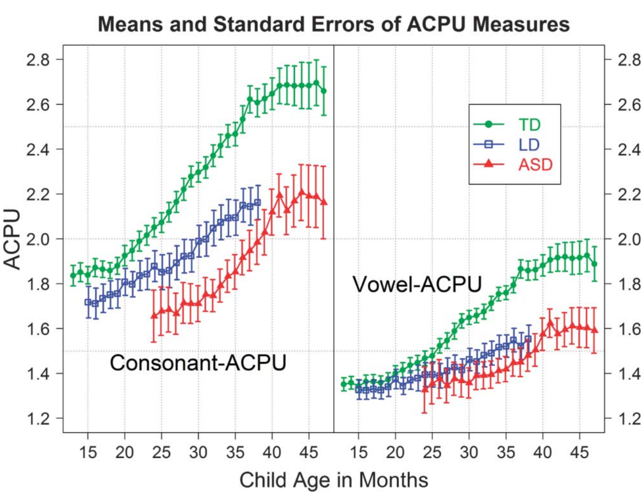


Table 5. Weighted correlations between ACPU measures and child chronological month-age.

Measure	TD sample (n = 106)	LD sample (n = 49)	ASD sample (n = 71)
Consonant-ACPU	0.63***, †††, ‡‡‡	0.32*, †, ‡‡	0.32*, †, ‡
Vowel-ACPU	0.58***, †††, ‡‡‡	0.19	0.25*, †, ‡

Note. The statistical significances of the weighted correlations were obtained using two methods: theoretical estimation and bootstrap simulation, shown in Table 6. Theoretical estimation is based on certain assumptions, which may or may not be true for a specific sample, whereas bootstrap simulation is a nonparametric approach that resamples the distribution of the statistics of interest, which cannot be repeated exactly and needs more computation. The two methods are complementary to some extent. For a rigorous analysis, higher p values of the two methods were chosen as combined final p values. As shown, the two methods provided consistent statistical significances most of the time.

*p < .05. ***p < .001 (combined final p value).

†p < .05. †††p < .001 (p value using theoretical estimation).

‡p < .05. ‡‡p < .01. ‡‡‡p < .001 (p value using bootstrap simulation).

z scores for ACPU measures in each recording relative to age means and age variances in the TD group and then averaged the z scores across each child's recordings to obtain child z scores. The LD and ASD samples were also age-standardized to the TD group. With age-standardized child z scores, the group difference analysis was straightforward using one-way ANOVA to test the significance of group difference; Welch two-sample t tests were conducted for each pair of TD, LD, and ASD groups to see which groups differed.

Figure 7 illustrates the distribution of child z-scores in the three groups using boxplots for comparison. In our analysis, Consonant-ACPU and Vowel-ACPU (z-scores) were significantly higher in the TD group than in the LD and ASD groups; see Table 7. Consonant-ACPU, but not Vowel-ACPU, is only significantly higher (although with relatively small difference) in the LD than ASD group.

Summary

The results of this case study were consistent with our prestudy predictions, which were based on published findings using conventional approaches. In short, machine-based ACPU measures demonstrated significant group differences among the three child groups studied and showed that child developmental trends were significantly correlated with age. Our analysis suggests that the new automated approach, which is objective, easy to use, and cost effective, holds great promise because it can accommodate large numbers of children with fewer resources than used in conventional approaches.

Table 6. Mean group differences of weighted correlations using bootstrap simulation.

Measure	TD–ASD	TD–LD	LD–ASD
Consonant-ACPU	0.31**	0.31***	0.01
Vowel-ACPU	0.33**	0.39***	–0.06

**p < .01. ***p < .001 (combined final p value).

Discussion

Phonetic development in young children serves as a foundation for language acquisition, a sufficiently comprehensive analysis of which could prove clinically valuable. Yet, existing research in this area has been limited by high costs, time investment, and the level of expertise needed to collect and transcribe phonetic development data on a large scale in naturalistic settings using conventional methods. We explored a novel automated approach that provides an opportunity to alleviate the burden of conventional methods.

The first stage of automated data processing in the new approach was sound segmentation, which was validated by 70–82 hr of human-transcribed audio data. Clear key child sound segments (utterances) were identified with more than 72% accuracy.

The second stage further processed the identified child utterances using the open-source Sphinx phone recognizer and produced estimates of Consonant-ACPU and Vowel-ACPU. The human-transcribed data set showed that machine estimates were significantly correlated with human-transcribed ACPUs (greater than 0.82), but were underestimated. By including machine-recognized nonspeech units, the total machine count for consonants, vowels, and nonspeech units exceeded the total human count. This finding suggests that the machine algorithm considered more details of child utterances than did human transcribers, and these were based on phonetic acoustic models without any lexicon and grammar knowledge and were sensitive to affecting factors, such as recording audio quality; environment

Figure 7. Box plot of child z scores of ACPU measures. The box plot demonstrates data distribution by showing median values, maximum and minimum values, and lower and upper quartiles with outliers, if any.

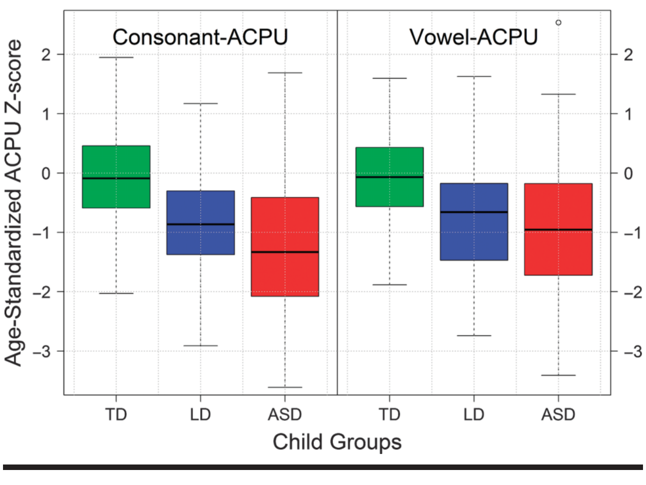


Table 7. Diagnostic group comparisons with age-standardized *z* scores of ACPUs.

Measure	Mean and standard deviation			ANOVA: $F(2, 223)$	t values of two-way comparison		
	TD (n = 106)	LD (n = 49)	ASD (n = 71)		TD-ASD	TD-LD	LD-ASD
Consonant-ACPU	–0.05 (0.75)	–0.83 (0.84)	–1.22 (1.16)	37.4***	7.55***	5.60***	2.16*
Vowel-ACPU	–0.08 (0.75)	–0.73 (0.89)	–1.00 (1.24)	21.3***	5.59***	4.48***	1.35

Note. ANOVA = analysis of variance.
*p < .05. ***p < .001.

noise; and not-well-formed pronunciations, with the tendency of recognizing "contaminated" or "distorted" sound units as nonspeech units. In contrast, human transcribers perceived words using contextual information and knowledge of lexicon and grammar, so could therefore filter out some of the affecting factors and ignore certain details. This helps to explain the higher overall machine count, but fewer machine-recognized consonants and vowels. In general, the high correlation between machine and human ACPUs indicates that when human ACPUs are relatively higher or lower for a child among some children or compared among this child's different ages, the corresponding machine ACPUs would also be higher or lower relative to the machine ACPUs of other children or different ages of the same child. Relative values are more important than absolute values in depicting relative comparisons.

The effectiveness of the new approach was exemplified by its application to the study of group differences using Consonant-ACPU and Vowel-ACPU and child developmental trends. Our study showed that the Consonant-ACPU in the TD group was significantly correlated with child age (as high as 0.63), whereas such correlations in the LD and ASD groups were significantly weaker (0.32). When the three child groups were compared in terms of age-standardized *z* scores, children with TD produced significantly more consonants than did children with LD, who in turn produced significantly more consonants than did children with ASD. These results were consistent with those in the literature (Paul et al., 2011) and were essentially supported by the longitudinal analysis using the HLM method, as presented in the Appendix.

Child phonetic development in terms of vowel count has not often been reported in the literature. Paul et al. (2011), for instance, studied the counts for consonants and canonical syllables, but not individual vowel counts. Tager-Flusberg et al. (2009) recommended "expressive language benchmarks for children with ASD," but these included only consonant inventory and consonant-vowel combination as phonological variables without individual vowel variables. In the current study, however, we collected large samples with the opportunity to check vowel counts individually. The analyses showed that children with TD produced significantly more vowels per utterance as they grew older (0.58 correlates with child age), which was significantly higher when compared with those of children with LD and ASD. Nevertheless, vowel count per utterance (Vowel-ACPU) cannot be used to distinguish children with LD from children with ASD. These results were also essentially supported by the additional longitudinal analysis presented in the Appendix.

Distinguishing children with developmental problems from children with TD is an important goal. Equally important would be the ability to distinguish among different child developmental issues using this methodology. The differences observed between the children with LD and ASD in the current study are of particular interest. The LD group produced significantly more consonants per utterance (Consonant-ACPU) than did the ASD group, which may suggest that phonetic development in the ASD group may be even more delayed than that in the LD group or that the "sound quality" of the phonetic units produced by children with ASD may not be as good as that of children with LD. This difference was captured by the adult acoustic phonetic models used, which resulted in fewer consonants being recognized by machine in children with ASD. Further investigation in this area, including more detailed analyses with matched samples in language development and further grouping of the ASD group into high- and low-functioning subgroups, are clearly needed and warranted.

Given the strong developmental trends and patterns of differences among groups shown here, we conclude that the automated approach used in this study can efficiently and effectively produce meaningful estimates related to phonetic development. These results suggest great potential for this approach to be incorporated into standard scientific research and clinical practices.

Clinically, the automated approach could be helpful in the efforts of early screening, diagnosing, and intervention of development delay and disorders that may otherwise be cost prohibitive (e.g., in rural communities and underserved populations).

Aside from cost factors, the vast data samples that one is able to collect using the automated approach (literally millions of utterances in this study) also provide a new degree of robustness (i.e., resistance to the confounding effects of behavior variation and measurement errors).

Despite the promising results observed in the current study, several limitations to the current approach are acknowledged. First, it is not clear—even with our large quantity of sampling—how many daylong recordings over what duration of time would be necessary to establish a reliable and consistent estimate of any vocal measure, nor is it clear to what extent the number of recordings affects the robustness of the acquired data. More detailed phonetic categorization may be needed to allow for a more comprehensive phonetic analysis and the study of additional acoustic, phonetic, and prosodic features.

Although computational approaches for gathering and analyzing data are widely applied in physics, chemistry, biology, the social sciences, and other areas of research, the fundamental paradigm in child language development research—and especially phonetic development—remains based primarily on subjective human efforts. Results from the current study indicate that automated, computational, and objective approaches are efficient and effective tools for obtaining and analyzing large samples of child phonetic data and providing new windows into the study of early child development.

Acknowledgment

We gratefully acknowledge Terrance D. Paul for his conception of the LENA tool and for personally funding and directing its development, as well as the development of the LENA Natural Language Corpus.

References

Chernick, M. R., & LaBudde, R. A. (2011). *An introduction to bootstrap methods with applications to R*. New York, NY: Wiley.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Davis, B. L., MacNeilage, P. F., & Matyear, C. L. (2002). Acquisition of serial complexity in speech production: A comparison of phonetic and phonological approaches to first word production. *Phonetica, 59,* 75–107.

Dodd, B., & McIntosh, B. (2010). Two-year-old phonology: Impact of input, motor and cognitive abilities on development. *Journal of Child Language, 37,* 1027–1046. doi:10.1017/S0305000909990171

Dyson, A. T. (1988). Phonetic inventories of 2- and 3-year-old children. *Journal of Speech and Hearing Disorders, 53,* 89–93.

Ertmer, D. J. (2001). Emergence of a vowel system in a young cochlear implant recipient. *Journal of Speech, Language, and Hearing Research, 44,* 803–813.

Ferguson, C. A., & Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language, 51,* 419–439.

Grizzle, K. L., & Simms, M. D. (2005). Early language development and language learning disabilities. *Pediatrics in Review, 26,* 274–283. doi:10.1542/pir.26-8-274

Ingram, D. (2002). The measurement of whole-word productions. *Journal of Child Language, 29,* 713–733.

Ingram, D., & Ingram, K. D. (2001). A whole-word approach to phonological analysis and intervention. *Language, Speech, and Hearing Services in Schools, 32,* 271–283. doi:10.1044/0161-1461(2001/024)

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience, 5,* 831–843. doi:10.1038/nrn1533

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *The Journal of the Acoustical Society of America, 100,* 2425–2438. doi:10.1121/1.417951

McCune, L., & Vihman, M. M. (2001). Early phonetic and lexical development: A productivity approach. *Journal of Speech, Language, and Hearing Research, 44,* 670–684. doi:10.1044/1092-4388(2001/054)

Nathani, S., Ertmer, D. J., & Stark, R. E. (2006). Assessing vocal development in infants and toddlers. *Clinical Linguistics & Phonetics, 20,* 351–369. doi:10.1080/02699200500211451

Oller, D. K. (1980). The emergence of the sounds of speech in infancy. In G. Yeni-Komshian, J. Kavanagh, & C. Ferguson (Eds.), *Child phonology* (pp. 93–112). New York, NY: Academic Press.

Oller, D. K. (2000). *The emergence of the speech capacity*. Mahwah, NJ: Erlbaum.

Oller, D. K., Eilers, R. E., Bull, D. H., & Carney, A. E. (1985). Pre-speech vocalizations of a deaf infant: A comparison with normal metaphonological development. *Journal of Speech and Hearing Research, 28,* 47–63.

Oller, D. K., Eilers, R. E., Steffens, M. L., Lynch, M. P., & Urbano, R. (1994). Speech-like vocalization in infancy: An evaluation of potential risk factors. *Journal of Child Language, 21,* 33–58.

Oller, D. K., Niyogi, P., Gray, S., Richards, J. A., Gilkerson, J., Xu, D., . . . Warren, S. (2010). Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. *Proceeding of the National Academy of Sciences of the United States of America, 107,* 13354–13359.

Paul, R., Fuerst, Y., Ramsay, G., Chawarska, K., & Klin, A. (2011). Out of the mouths of babes: Vocal production in infant siblings of children with ASD. *Journal of Child Psychology and Psychiatry, 52,* 588–598.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech, 46,* 115–154.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2008). *Linear and nonlinear mixed effects models* (R Package Version 3.1-89). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

R Development Core Team. (2011). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Rescorla, L., & Bernstein-Ratner, N. A. (1996). Phonetic profiles of toddlers with specific expressive language impairment (SLI-E). *Journal of Speech and Hearing Research, 39,* 153–165.

Saaristo-Helin, K. (2011). *Phonological mean length of utterance as an indicator of typical and impaired phonological development in children acquiring Finnish* (Academic dissertation, University of Helsinki, Finland). Retrieved from https://helda.helsinki.fi/handle/10138/28172?show=full

Saaristo-Helin, K., Kunnari, S, & Savinainen-Makkonen, T. (2011). Phonological development in children learning Finnish: A review. *First Language, 31,* 342–363.

Saaristo-Helin, K., Savinainen-Makkonen, T., & Kunnari, S. (2006). The phonological mean length of utterance: Methodological challenges from a crosslinguistic perspective. *Journal of Child Language, 33,* 179–190.

Serkhane, J. E., Schwartz, J. L., Boë, L. J., Davis, B. L., & Matyear, C. L. (2007). Infants' vocalizations analyzed with an articulatory model: A preliminary report. *Journal of Phonetics, 35,* 321–340.

Seymore, K., Chen, S., Doh, S., Eskenazi, M., Gouvea, E., Raj, B., . . . Thayer, E. (1998). The 1997 CMU Sphinx-3 English broadcast news transcription system. *Proceedings of the DARPA Workshop on Broadcast News Understanding Systems (BNTUW–98),* Lansdowne, VA. Retrieved from http://www.itl.nist.gov/iad/mig/publications/proceedings/darpa98/index.htm

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis*. New York, NY: Oxford University Press.

Stark, R. E. (1980). Prespeech segmental feature development. In G. Yeni-Komshian, J. Kavanagh, & C. Ferguson (Eds.), *Child phonology* (pp. 73–92). New York, NY: Academic Press.

Stark, R. E. (1983). Phonatory development in young normally hearing and hearing-impaired children. In I. Hochberg, H. Levitt, & M. J. Osberger (Eds.), *Speech of the hearing impaired: Research, training, and personnel preparation* (pp. 297–312). Baltimore, MD: University Park Press.

Stoel-Gammon, C. (1985). Phonetic inventories, 15–24 months: A longitudinal study. *Journal of Speech and Hearing Research, 28,* 505–512.

Stoel-Gammon, C. (1988). Prelinguistic vocalizations of hearing-impaired and normally hearing subjects: A comparison of consonantal inventories. *Journal of Speech and Hearing Disorders, 53,* 302–315.

Stoel-Gammon, C. (1991). Normal and disordered phonology in two-year-olds. *Topics in Language Disorders, 11,* 21–32.

Stoel-Gammon, C. (2010). The word complexity measure: Description and application to developmental phonology and disorders. *Clinical Linguistics & Phonetics, 24,* 271–282. doi:10.3109/02699200903581059

Tager-Flusberg, H., Rogers, S., Cooper, J., Landa, R., Lord, C., Paul, R., . . . Yoder, P. (2009). Defining spoken language benchmarks and selecting measures of expressive language development for young children with autism spectrum disorders. *Journal of Speech, Language, and Hearing Research, 52,* 643–652.

Thomas, S. C. M. (2010). Language acquisition in developmental disorders. In M. Kail & M. Hickmann (Eds.), *Language acquisition across linguistic and cognitive systems* (pp. 67–90). Amsterdam, the Netherlands: John Benjamins.

Vihman, M. M. (1996). *Phonological development: The origins of language in the child*. Cambridge, MA: Blackwell.

Vihman, M. M., Keren-Portnoy, T., Bidgood, A., McGillion, M., & Whitaker, C. (2013). Late talking toddlers: Relating early phonological development to later language advance, *York Papers in Linguistics*. Retrieved from http://www.york.ac.uk/language/ypl/current.html

Williams, A. L., & Elbert, M. (2003). A prospective longitudinal study of phonological development in late talkers. *Language, Speech, and Hearing Services in Schools, 34,* 138–153.

Xu, D., Gilkerson, J., Richards, J., Yapanel, U., & Gray, S. (2009). Child vocalization composition as discriminant information for automatic autism detection. In *Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society,* 2518–2522, September 3–6, 2009, Minneapolis, MN, doi:10.1109/IEMBS.2009.5334846

Xu, D., Yapanel, U., Gray, S., Gilkerson, J., Richards, J. A., & Hansen, J. (2008, October). *Signal processing for young child speech language development*. Paper presented at the First Workshop on Child, Computer and Interaction, Chania, Crete, Greece. Retrieved from http://www.lenafoundation.org/DownloadFile.aspx/pdf/SignalProcessing_ChildSpeech

Appendix

Longitudinal Analysis With Hierarchical Linear Modeling in Three Child Groups With ACPU Measures

Some analyses in the main text were based on nonparametric methods. Comparatively, model-based parametric analyses may enable better data summary and provide differing views. We also conducted a longitudinal analysis using hierarchical linear modeling (HLM), the results of which are described here to offer additional (but essentially consistent) evidence to supplement the findings in the main text.

We considered two issues before starting the analysis with HLM or linear models. First, we confirmed the linearity or nonlinearity of the data. Based on the descriptive statistics of age means and age standard errors of ACPU measures (see Figure 6), we observed that the development curves were not always linear in large age ranges but were linear in some smaller age ranges. Considering both data linearity and availability in different age ranges, we chose the following for analysis and comparison: children with typical development (TD) ages 15–38 months, children with language delay (LD) ages 15–38 months, and children with autism spectrum disorder (ASD) ages 16–38 months.

Second, longitudinal data were lacking for the ASD group. As shown in Figure 5, most of the children had recording(s) in only 1 month, and only a few children had recordings spanning over only 3 or 4 months. According to Singer and Willett (2003), the slope or rate of change cannot be reliably estimated from the data with short time span, and our analysis verified their assertion. Therefore, the ASD group was treated as cross-sectional data and modeled with weighted least squares (WLS), using the weight of *1/n* for each of the recordings of a given child, if the child had *n* recordings. In so doing, each child had an overall weight of 1 and was treated equally in the modeling process. All analyses used the R programming language (R Development Core Team, 2011) and the "nlme" package (Pinheiro et al, 2008) for the longitudinal analysis with HLM.

Given that within-group variations for each child's linear model (i.e., the intercept and the slope of the model) could differ between two groups, TD and LD were not modeled together to share the same distribution for within-group intercept and slope residuals but were instead modeled separately.

The means and standard errors of the intercepts and slopes of the linear models for the three groups were individually estimated and were further two-way compared using the Welch two-sample *t* test. The results are shown in Table A1, and the linear models are illustrated in Figure A1.

The analyses conducted here were mixed in type and were not uniform; however, the results can be used as a reference for comparison with the previous analyses. For the TD and LD groups with longitudinal data, a standard longitudinal analysis would model both groups together, sharing distributions of within-group variations. To further explore and verify the consistency of the results, the two groups were modeled together using the HLM method (see Table A2).

Table A2 shows that the results of the "standard" longitudinal analysis using HLM are consistent with those in Table A1, which partially supports the mixed analyses scheme used in Table A1. We focus on the results in Table A1 and the illustration in Figure A1 to present a brief summary here. First, we observed the following with regard to the intercepts in the linear models or the ACPU measures at 30 months: (a) all were significantly above 0, (b) ACPUs in the TD group were all significantly higher than those in the other two groups, and (c) the LD group was significantly higher than the ASD group in Consonant-ACPU but not Vowel-ACPU. Second, we observed the following with regard to the slopes or rates of change in the linear models: (a) only the TD and LD groups were significantly higher than 0, (b) the rate of change of the ASD group did not differ significantly from 0 for both ACPU measures, and (c) group differences were significant only when TD was compared with ASD for both ACPUs or Vowel-ACPU was compared in the TD and LD groups. These results provide a perspective from linear growth models and are essentially consistent with findings in the analyses of the main text.

Table A1. ACPU HLM and WLS modeling results: Diagnostic group statistics and comparisons.

	Mean and standard error			*t* values of two-way comparison		
Measure	**TD (618, 90)[a]**	**LD (284, 44)[a]**	**ASD (93, 42)[a]**	**TD-ASD**	**TD-LD**	**LD-ASD**
Consonant-ACPU intercept at 30 months	2.25*** (.040)	1.95*** (.064)	1.72*** (.039)	5.59***	4.11***	2.23*
Consonant-ACPU slope: Change/10 months	.363*** (.037)	.232*** (.069)	.069 (.075)	3.12**	1.81	1.34
Vowel-ACPU intercept at 30 months	1.61*** (.026)	1.44*** (.037)	1.35*** (.032)	3.99***	3.55***	1.41
Vowel-ACPU slope: Change/10 months	.227*** (.026)	.114* (.048)	.063 (.062)	2.46*	2.25*	0.59

Note. The TD and LD groups were analyzed using HLM. ASD was modeled as a standard cross-sectional analysis using WLS. Group means and standard errors for both intercepts and slopes were thus estimated individually using their respective methods. ACPU = average count per utterance; HLM = hierarchical linear modeling; WLS = weighted least squares; TD = typical development; LD = language delay; ASD = autism spectrum disorder.

[a] The first number is the number of recordings; the second number is the number of children.

$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Figure A1. Group linear models for children with TD, LD, and ASD. The intercepts and slopes of the linear models shown here are further described in Table A1. The solid lines represent the results using the hierarchical linear modeling method for both the TD and LD groups, whereas the dashed lines represent the results using the weighted least squares method for the ASD group (as explained in Table A1). The dots represent the intercepts at 30 months for each group.

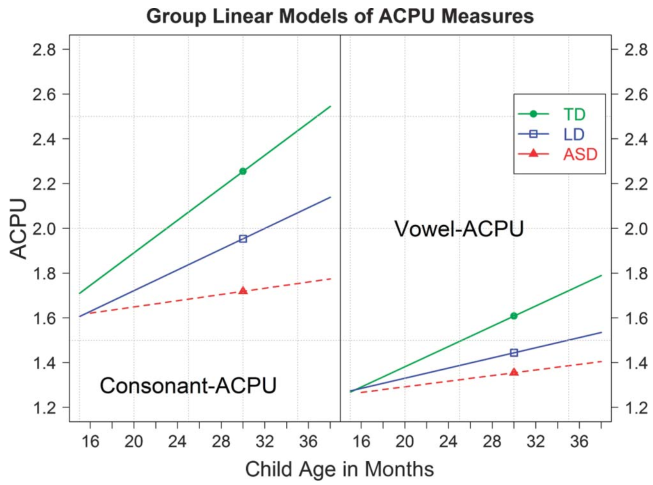


Table A2. Longitudinal analysis of ACPUs of TD and LD groups using HLM.

	Mean and standard error		
Measure	**TD (618, 90)[a]**	**LD (284, 44)[a]**	**TD – LD (Mean difference between TD & LD)**
Consonant-ACPU intercept at 30 months	2.25*** (.041)	1.97*** (.058)	0.28*** (.071)
Consonant-ACPU slope: Change/10 months	.361*** (.039)	.261*** (.066)	.100 (.076)
Vowel-ACPU intercept at 30 months	1.61*** (.027)	1.44*** (.037)	0.17*** (.045)
Vowel-ACPU slope: Change/10 months	.225*** (.026)	.108* (.047)	.117* (.053)

Note. The ASD group was not included in this analysis because it lacked longitudinal data; therefore, the slopes of each child model could not be reliably estimated. TD and LD groups were analyzed together using HLM as a standard longitudinal analysis. The results were consistent with and partially supported the corresponding results in Table A1.

[a] The first number is the number of recordings; the second number is the number of children.

*$p < .05$. ***$p < .001$.