

## Measures

### Behavioral Measures

**Proportion Attempted (Index 1 in Supplemental Table 1).** While most participants attempted all of the verbal tasks, some children did not. Empty cells for measures for these productions would result in less-willing participants self-selecting out of the analyses, as missing data decreased the amount of information contributed by these individuals. Accordingly, the overall proportion of tasks attempted by each participant was entered into the statistical model.

**Proportion With Phonemes Correct (Index 2 in Supplemental Table 1).** The overall proportion of verbal productions transcribed that matched the intended target was reported for each participant to provide a measure of phonemic accuracy. Data from all attempted verbal productions (i.e., trochees, iambs, and nonwords) were collapsed to one input measure for each participant.

**Proportion of Trochees With Correct Stress (Index 3 in Supplemental Table 1).** The proportion of attempted trochees produced with correct stress was included as input to the subgroup discovery algorithm because ability to produce accurate lexical stress has been proposed to be a marker for suspected childhood apraxia of speech (Shriberg, Campbell, et al., 2003).

**Proportion Iambs With Correct Stress (Index 4 in Supplemental Table 1).** The proportion of attempted iambs produced with correct stress was included in analyses because ability to produce accurate lexical stress has been proposed to be a marker for suspected childhood apraxia of speech (Shriberg, Campbell, et al., 2003).

**Proportion Used (Index 5 in Supplemental Table 1).** Of all the productions attempted by a participant, this value reports the proportion that was produced with both phonemic and lexical stress accuracy and that also had usable kinematic trajectories that were free of movement artifact or markers obscured by hand motion. Often, spurious movements were produced by participants who were frustrated by a particular task; this would substantially reduce the amount of usable data generated. While productions with accurate phonemics and stress could be used for acoustic measures, they could not be included in the kinematic-level measures.

### **Acoustic Measures**

**Mean Acoustic Area Ratio—Trochees and iambs (two measures; Indices 6 and 7 in Supplemental Table 1).** For the lexical stress task, the primary acoustic analysis used was the Acoustic Area Ratio (adapted from Xie, Andreae, Zhang, & Warren, 2004), which quantified the relative energy of the first syllable to the second syllable. For each syllable, the mean amplitude was multiplied by the vowel duration to provide a measure of acoustic “area” (i.e., under the rectified and smoothed acoustic signal). Because the audio signals were not calibrated, amplitude was measured using the root-mean-square (RMS) of the signal in 20-ms windows. The mean amplitude for each participant and task was expressed as an average of these values. The Acoustic Area Ratio is the quotient of acoustic area of the first syllable and the second syllable; values greater than 1.0 indicate that primary stress was on the first syllable, those less than 1.0 indicate that primary stress was on the second syllable. The measure is similar to a metric found to be useful in the identification of subgroups of children with suspected apraxia of speech, the Lexical Stress Ratio (Shriberg, Campbell, et al., 2003).

**Coefficient of Variation of Acoustic Area Ratio—Trochees and iambs (two measures; Indices 8 and 9 in Supplemental Table 1).** Variability of the acoustic marking of lexical stress was evaluated using a normalized measure of variability, the coefficient of variation. Coefficient of variation is a unit-free measure of dispersion, defined as the quotient of the standard deviation and the mean. This measure allows a comparison of the degree of variation from one measure to another, even when the means and units of measure are drastically different.

**Word Duration—two- and three-syllable nonwords (two measures; Indices 10 and 11 in Supplemental Table 1).** Word duration, in seconds, was measured for the two- and three-syllable nonwords from the burst of the /b/ or the initial glottal pulse of the /m/ to the final glottal pulse of the second /a/. As a convention, the glottal offset was defined as the final negative-going zero crossing in the periodic signal associated with the vowel. Measures from the two-syllable nonwords (*bama* and *bada*) were collapsed, as were measures from the three-syllable nonwords (*bamana* and *manaba*). These measures were included because the duration of speech events has been shown to be a potential marker for some subgroups of speech delay (Shriberg, Flipsen, Kwiatkowski, & McSweeny, 2003).

**Coefficient of Variation of Word Duration—two- and three-syllable nonwords (Indices 12 and 13 in Supplemental Table 1).** These variables measured the consistency of word duration produced by each participant and were included in the model to identify potential subgroups of children who were more variable in speech rate.

### **Kinematic Measures**

**Mean Maximum Displacement—upper lip, lower lip, and jaw displacements were obtained for both verbal (Indices 14–16) and nonverbal (Indices 38 and 39) tasks.** Maximum

displacement is included in most studies of speech kinematics and was included in this analysis model with the assumption that it would help to identify children with aberrant speech movement characteristics.

The maximum vertical displacement of the first syllable of all productions was measured for the upper lip, lower lip, and jaw. The measure was made algorithmically using custom scripts in MATLAB that identified the local minimum (jaw and lower lip) or maximum (upper lip) adjacent to the first vowel onset, using the zero crossing in the velocity profile to pinpoint the precise location. The lower lip displacement record resulted from subtracting the displacement record of the jaw from the lower lip, point by point. Measures were collapsed across productions in order to minimize the number of missing cells in the analysis model. For instance, many children produced no iambs correctly. Collapsing across trochees and iambs ensured that every child had at least one kinematic measure in the analysis model. The risk of collapsing these data, of course, was that the smaller displacement typically observed on the first syllable of an iambic production relative to trochaic productions may affect the overall average. Since children in Group B produced so few iambs, this would result in larger average maximum displacements. Maximum displacement of the articulators for the second syllable was not measured because there was often not a reliable zero crossing in the velocity profile for the second syllable. This was due to the fact that the words were produced in isolation.

**Coefficient of Variation for Maximum Displacement—upper lip, lower lip, jaw for both verbal (Indices 17–19) and nonverbal tasks (Indices 40 and 41).** The variability of first-syllable maximum displacement was measured using the coefficient of variation and was included in the analysis model because of the possibility that increased speech movement variability may characterize a subgroup of children with differences in speech motor control.

**Spatiotemporal Index—trochees (Indices 20, 23, 26), iambs (Indices 21, 24, 27), three-syllable nonwords (Indices 22, 25, 28), and nonverbal tasks (Indices 42, 43); upper lip, lower lip, and jaw (11 measures).** Measures of movement stability were made for the whole-word movement records for trochaic, iambic, and three-syllable nonword productions for each single articulator. The measure was the spatiotemporal index (STI) developed by Smith et al. (1995). The STI uses time- and amplitude-normalized records across repetitions of each whole-word production. Amplitude normalization of individual trajectories was accomplished by subtracting the mean from each displacement record and then dividing each by its standard deviation. Linear temporal normalization was achieved by interpolating each amplitude normalized trajectory to 1,000 points using a cubic spline fitting algorithm in MATLAB. The standard deviation for all repetitions of successive times in the aligned normalized signals was calculated every 20 samples (2% intervals, totaling 50 standard deviations); the sum of these 50 standard deviations yielded the STI. The STI has been demonstrated to be sensitive to differences in performance attributable to specific diagnoses (e.g., SLI; Goffman, 1999). STI of coordinative measures, such as lip aperture, were made but not included in the analysis model to attempt to minimize the number of included measures and because they were highly correlated to the other STI measures (e.g., Pearson's  $r$  for jaw and lip aperture STI for trochaic productions = 0.671,  $p < .0001$ ).

For both the lexical stress task and the three-syllable nonword repetition task, STI was calculated for the upper lip, lower lip, and jaw markers' vertical displacements during each target production; only correct repetitions were included in these analyses (see "Proportion Used" above).

A variant of the STI was used for the nonverbal tasks, the cyclic spatiotemporal index (cSTI), which quantified the stability of repeated individual movement cycles (van Lieshout & Moussa, 2000). Calculation of the cSTI used the parsed open–close cycles from the jaw vertical displacement records, which were subsequently normalized and aligned by start and end times. The sum of the 50 standard deviations calculated at every 20th sample constitutes the cSTI.

**Convergence Index—trochees, iambs, and three-syllable nonwords; upper lip, lower lip and jaw (Indices 29–37 in Supplemental Table 1; nine measures).** A Convergence Index (CI; Goffman, Gerken, & Lucchesi, 2007) was also calculated; the CI is based on all attempted productions, including those that contained phonemic and lexical stress errors, as well as those produced accurately. The CI quantifies the stability of underlying movement patterns for the target productions, without respect to accuracy. The intention of including this measure was to provide a measure of kinematic stability, even for children who produced few, if any, repetitions with phonemic or lexical stress accuracy.

**PEPPER Measures (Indices 44–51 in Supplemental Table 1).** Programs to Examine Phonetic and Phonological Evaluation Records (PEPPER; Shriberg et al., 2001) was used to generate error profiles for each child from a 15-minute conversational speech sample. The PEPPER software environment was used (a) to classify participants' speech status and (b) to produce percentage scores on all of the competence variables (see below). The program yielded a number of speech performance metrics that were included as input to the classification model. Although these metrics are somewhat numerous, they are not redundant; each value provides an indication of kwiatperformance in a unique speech production domain that may have plausibly emerged from SUBARP.

**PCC:** Percentage of Consonants Correct; measure of phonetic accuracy, designed to quantify common and uncommon speech sound distortions, as systematized in the Appendix of Shriberg et al. (1997).

**PCCR:** Percentage of Consonants Correct, Revised; this index, which does not count distortions (allophonic variation) as errors, provides a measure of phonemic accuracy. This measure provides greater distinction between children with normal speech acquisition and those with speech delay in life span reference data. It reflects the acquisition of the phonemes in the participants' ambient community (Shriberg et al., 1997).

**PVC:** Percentage of Vowels Correct.

**PVCR:** Percentage of Vowels Correct, Revised; parallel to PCCR; does not count distortions as errors.

**PPC:** Percentage of Phonemes Correct.

**PPCR:** Percentage of Phonemes Correct, Revised; does not count distortions as errors.

**II:** Intelligibility Index; proportion of unintelligible words to total words produced, expressed as a percentage.

**AWU:** Average Words per Utterance; an estimate of the child's expressive language function and verbal productivity, which varies greatly among typically speaking children in any given continuous speech sample. AWU correlates greater than 90% with mean length of utterance in preschool-age children (Shriberg & Kwiatkowski, 1994). This output measure was used to index general lexical productivity, not grammar.

## References

- Goffman, L. (1999). Prosodic influences on speech production in children with specific language impairment and speech deficits: Kinematic, acoustic, and transcription evidence. *Journal of Speech, Language, and Hearing Research, 42*, 1499–1517.
- Goffman, L., Gerken, L., & Lucchesi, J. (2007). Relations between segmental and motor variability in prosodically complex nonword sequences. *Journal of Speech, Language, and Hearing Research, 50*, 444–458.
- Shriberg, L. D., Allen, C. T., McSweeney, J. L., & Wilson, D. L. (2001). *PEPPER: Programs to examine phonetic and phonologic evaluation records*. Madison: Waisman Center, University of Wisconsin—Madison.
- Shriberg, L. D., Austin, D., Lewis, B. A., McSweeney, J. L., & Wilson, D. L. (1997). The percentage of consonants correct (PCC) metric: Extensions and reliability data. *Journal of Speech, Language, and Hearing Research, 40*, 708–722.
- Shriberg, L. D., Campbell, T. F., Karlsson, H. B., Brown, R. L., McSweeney, J. L., & Nadler, C. J. (2003). A diagnostic marker for childhood apraxia of speech: The lexical stress ratio. *Clinical Linguistics & Phonetics, 17*, 549–574.
- Shriberg, L. D., Flipsen, P., Jr., Kwiatkowski, J., & McSweeney, J. L. (2003). A diagnostic marker for speech delay associated with otitis media with effusion: The intelligibility-speech gap. *Clinical Linguistics & Phonetics, 17*, 507–528.
- Shriberg, L. D., & Kwiatkowski, J. (1994). Developmental phonological disorders I: A clinical profile. *Journal of Speech, Language, and Hearing Research, 37*, 1100–1126.



Smith, A., Goffman, L., Zelaznik, H. N., Ying, G., & McGillem, C. (1995). Spatiotemporal stability and patterning of speech movement sequences. *Experimental Brain Research*, 104, 493–501.

Van Lieshout, P. H. H. M., & Moussa, W. (2000). The assessment of speech motor behavior using electromagnetic articulography. *The Phonetician*, 81, 9–22.

Xie, H., Andrae, P., Zhang, M., & Warren, P. (2004, January). Detecting stress in spoken English using decision trees and support vector machines. In *Proceedings of the second workshop on Australasian information security, Data Mining and Web Intelligence, and Software Internationalisation-Volume 32* (pp. 145–150). Sydney: Australian Computer Society.

**Table 1.** Descriptive statistics for each of the 53 continuous measures.

Index and measure	<i>M</i> ( <i>SD</i> )	Subgroup A	Subgroup B	Nonclassified	<i>t</i> (82)	<i>p</i>
		( <i>n</i> = 76): <i>M</i> ( <i>SD</i> )	( <i>n</i> = 10): <i>M</i> ( <i>SD</i> )	( <i>n</i> = 13): <i>M</i> ( <i>SD</i> )		
1. Proportion						
attempted	0.88 (0.23)	0.98 (0.05)	0.39 (0.19)	0.67 (0.26)	-22.57	<.0001
2. Proportion with						
phonetics						
correct	0.65 (0.23)	0.74 (0.14)	0.31 (0.18)	0.44 (0.30)	-8.65	<.0001
3. Proportion with						
stress correct—						
Trochees	0.80 (0.26)	0.88 (0.15)	0.47 (0.44)	0.63 (0.32)	-6.10	<.0001
4. Proportion with						
stress correct—						
Iambs	0.60 (0.32)	0.72 (0.23)	0.02 (0.05)	0.36 (0.31)	-9.66	<.0001
5. Proportion with						
no phonetic						
errors and good						
kinematics	0.54 (0.23)	0.63 (0.17)	0.24 (0.16)	0.28 (0.21)	-6.88	<.0001

6. Mean acoustic						
area ratio—						
Trochees	1.80 (0.83)	1.87 (0.91)	1.55 (0.44)	1.61 (0.21)	-1.10	.28
7. Mean acoustic						
area ratio—						
Iambs	0.73 (0.36)	0.70 (0.24)	0.61 (0.17)	0.94 (0.76)	-1.17	.24
8. CV of acoustic						
area ratio –						
Trochees	42.70 (17.70)	42.31 (18.10)	39.09 (19.66)	47.70 (13.66)	-0.52	.60
9. CV of acoustic						
area ratio—						
Iambs	43.21 (20.04)	42.97 (17.88)	27.98 (15.49)	56.26 (26.65)	-2.52	0.014
10. 2-syllable word						
duration						
(seconds)	0.78 (0.08)	0.77 (0.08)	0.85 (0.09)	0.79 (0.06)	2.79	.007
11. 3-syllable						
nonword						
duration						
(seconds)	1.47 (0.14)	1.47 (0.14)	1.53 (0.15)	1.47 (0.14)	1.36	.179
12. CV of 2-syllable	11.27 (3.24)	11.15 (2.52)	12.22 (5.12)	11.25 (4.99)	1.08	.282

word duration						
13. CV of 3-syllable nonword duration	9.65 (5.11)	9.46 (5.04)	9.87 (5.30)	10.52 (5.68)	0.24	.813
14. Mean maximum displacement—UL (cm)	0.19 (0.07)	0.19 (0.07)	0.23 (0.67)	0.16 (0.06)	4.58	<.0001
15. Mean maximum displacement—LL (cm)	0.24 (0.09)	0.24 (0.09)	0.26 (0.10)	0.21 (0.09)	0.49	.623
16. Mean maximum displacement—Jaw (cm)	0.70 (0.24)	0.71 (0.23)	0.74 (0.32)	0.61 (0.22)	0.40	.689
17. CV of maximum displacement—UL <sup>a</sup>	43.49 (15.07)	40.02 (12.85)	57.30 (14.90)	52.65 (17.96)	3.92	<.0001
18. CV of maximum displacement—LL <sup>a</sup>	38.99 (10.61)	37.06 (7.86)	44.36 (15.34)	45.88 (15.82)	2.41	.018
19. CV of maximum displacement—Jaw	34.87 (10.35)	32.14 (7.64)	45.99 (13.38)	41.86 (12.86)	4.86	<.0001

displacement—						
Jaw						
20. UL STI—						
Trochees	29.16 (5.92)	28.80 (5.96)	30.94 (7.74)	29.83 (4.02)	1.03	.308
21. UL STI—Iambs <sup>a</sup>	29.20 (7.51)	28.81 (7.87)	30.25 (7.57)	30.59 (5.33)	0.55	.585
22. UL STI—3-						
syllable						
nonwords <sup>a</sup>	25.64 (7.10)	25.30 (7.65)	25.84 (6.54)	27.39 (3.31)	0.21	.832
LL STI—Trochees	31.11 (6.01)	31.21 (5.28)	32.19 (8.15)	29.72 (8.20)	0.52	.607
24. LL STI—Iambs <sup>a</sup>	30.59 (6.72)	30.83 (6.70)	31.22 (7.94)	28.71 (6.00)	0.17	.869
25. LL STI—3-						
syllable						
nonwords <sup>a</sup>	28.86 (6.59)	28.94 (6.49)	29.09 (7.39)	28.23 (7.05)	0.07	.948
Jaw STI—Trochees	22.17 (5.20)	22.18 (5.39)	24.09 (5.36)	20.69 (3.52)	1.05	.295
27. Jaw STI—						
Iambs <sup>a</sup>	20.52 (6.20)	20.66 (6.71)	21.40 (4.56)	19.06 (3.86)	0.34	.734
28. Jaw STI—3-						
syllable <sup>a</sup>						
nonwords	26.91 (7.21)	27.22 (7.65)	26.04 (7.04)	25.78 (4.40)	-0.46	.646

29. UL CI—						
Trochees	30.20 (6.37)	29.39 (6.04)	33.03 (7.84)	32.67 (6.21)	1.72	.089
30. UL CI—Iambs <sup>a</sup>	30.14 (7.35)	29.44 (7.47)	31.90 (7.57)	32.80 (6.03)	0.98	.331
31. UL CI—3-						
syllable						
nonwords <sup>a</sup>	33.10 (8.75)	32.35 (8.94)	33.23 (8.22)	37.30 (7.22)	0.30	.768
32. LL CI—						
Trochees	31.67 (5.76)	31.75 (4.91)	33.50 (6.14)	29.82 (9.17)	1.02	.309
33. LL CI—Iambs <sup>a</sup>	31.02 (6.73)	31.02 (6.44)	32.43 (5.94)	29.96 (8.98)	0.65	.515
34. LL CI—3-						
syllable						
nonwords <sup>a</sup>	33.94 (6.97)	34.41 (5.89)	33.53 (9.52)	31.54 (10.07)	-0.41	.683
35. Jaw CI—						
Trochees	22.78 (4.90)	22.43 (4.51)	25.81 (4.42)	22.43 (6.68)	2.22	.029
36. Jaw CI—Iambs <sup>a</sup>	21.05 (6.41)	20.94 (6.77)	22.55 (3.86)	20.51 (6.04)	0.73	.466
37. Jaw CI—3-						
syllable						
nonwords <sup>a</sup>	29.86 (7.02)	29.88 (6.66)	28.72 (5.58)	30.64 (9.97)	-0.53	.601
38. Jaw maximum	1.04 (0.42)	1.05 (0.40)	1.05 (0.20)	0.96 (0.67)	0.05	.963

displacement—						
vertical jaw						
oscillations						
(cm)						
39. Jaw maximum						
displacement—						
chewing (cm)	0.98 (0.23)	0.99 (0.22)	1.08 (0.24)	0.84 (0.26)	1.21	.23
40. CV of jaw						
maximum						
displacement—						
oscillations	39.6 (23.8)	37.8 (22.9)	31.9 (12.4)	56.4 (29.4)	−0.79	.432
41. CV of jaw						
maximum						
displacement—						
chewing	32.1 (8.5)	31.7 (8.5)	33.1 (4.3)	33.9 (11.1)	0.51	.612
42. cSTI—						
oscillation	14.81 (4.94)	14.41 (4.89)	16.40 (5.73)	15.92 (4.51)	1.18	.240
43. cSTI—chewing	26.85 (5.55)	26.59 (5.58)	26.80 (1.18)	28.40 (7.28)	0.52	.607
44. Percentage of						
consonants	71.96 (10.55)	71.99 (10.64)	70.34 (9.56)	73.06 (11.35)	−0.46	.644

correct <sup>a</sup>						
45. Percentage of consonants						
correct, revised	77.05 (10.27)	77.25 (77.25)	76.36 (7.79)	76.41 (10.34)	-0.26	.799
46. Percentage of vowels correct	91.99 (5.41)	91.93 (4.92)	91.06 (8.51)	93.08 (5.55)	-0.48	.636
47. Percentage of vowels correct, revised	94.05 (4.67)	93.97 (4.76)	94.16 (4.79)	94.39 (4.36)	0.12	.907
48. Percentage of phonemes correct <sup>a</sup>	80.00 (7.91)	79.99 (7.89)	78.69 (8.72)	81.11 (7.90)	-0.48	.630
49. Percentage of phonemes correct, revised	83.87 (7.58)	83.96 (7.87)	83.52 (6.38)	83.63 (7.28)	-0.17	.864
50. Intelligibility index	88.42 (11.28)	90.12 (10.00)	83.52 (6.38)	82.52 (15.71)	-1.94	.056
51. Average words per utterance <sup>a</sup>	2.70 (0.79)	2.74 (0.79)	2.81 (0.71)	2.38 (0.82)	0.28	.782



52. Age (months) <sup>a</sup>	46.02 (6.36)	46.70 (6.30)	42.50 (5.80)	44.85 (6.48)	-2.00	.049
53. Mother's education (years) <sup>a</sup>	14.69 (2.52)	14.57 (2.51)	15.70 (3.16)	14.57 (1.99)	1.29	0.2

---

*Note.* CI = confidence interval; cSTI = cyclic spatiotemporal index; LL = lower lip; STI = spatiotemporal index; UL = upper lip.

<sup>a</sup>Attribute only in SUBARP. Results from *t* test (*t* and *p* values) compare Subgroup A with Subgroup B.