JSLHR

Supplement Article

A Diagnostic Marker to Discriminate Childhood Apraxia of Speech From Speech Delay: IV. The Pause Marker Index

Lawrence D. Shriberg,^a Edythe A. Strand,^b Marios Fourakis,^a Kathy J. Jakielski,^c Sheryl D. Hall,^a Heather B. Karlsson,^a Heather L. Mabie,^a Jane L. McSweeny,^a Christie M. Tilkens,^a and David L. Wilson^a

Purpose: Three previous articles provided rationale, methods, and several forms of validity support for a diagnostic marker of childhood apraxia of speech (CAS), termed the pause marker (PM). Goals of the present article were to assess the validity and stability of the PM Index (PMI) to scale CAS severity. **Method:** PM scores and speech, prosody, and voice precision-stability data were obtained for participants with CAS in idiopathic, neurogenetic, and complex neurodevelopmental disorders; adult-onset apraxia of speech consequent to stroke and primary progressive apraxia; and idiopathic speech delay. Three studies were completed including criterion and concurrent validity

four-article series (Shriberg et al., 2017a, 2017b, 2017c, and the present article) to develop and assess the theoretical and clinical utility of a singlesign, behavioral marker of childhood apraxia of speech (CAS) poses the following three questions:

- 1. Do findings from construct and criterion validity studies support the diagnostic accuracy of a behavioral marker, the pause marker (PM), to discriminate CAS from speech delay (SD)?
- 2. Do findings from the PM and other measures support the hypothesis of core representational and transcoding deficits in CAS, and is the PM theoretically coherent with those deficits?

^bDepartment of Neurology, Mayo Clinic-Rochester, NY

[°]Department of Communication Sciences and Disorders, Augustana College, Rock Island, IL

Revision received August 2, 2016 Accepted August 21, 2016

https://doi.org/10.1044/2016_JSLHR-S-16-0149

studies of the PMI and a temporal stability study of the PMI using retrospective case studies. **Results:** PM scores were significantly correlated with other signs of CAS precision and stability. The best fit of the distribution of PM scores to index CAS severity was obtained by dividing scores into 4 ordinal severity classifications: mild, mild-moderate, moderate-severe, and severe. Severity findings for the 4 classifications and retrospective longitudinal findings from 8 participants with CAS supported the validity and stability of the PMI. **Conclusion:** Findings support research and clinical use of the PMI to scale the severity of CAS.

3. Do findings from cross-sectional and retrospective longitudinal case studies support an ordinal scale of PM scores, the Pause Marker Index (PMI), to quantify the severity of CAS for research and clinical applications?

The short-term goal of the PM is to provide a behavioral marker of CAS for research and clinical practice. A longer term goal of the PM is to provide quantitative inclusionary and exclusionary criteria to identify participants who are true positive for CAS in studies to identify and validate a biomarker of CAS. The present article addresses both goals. As proposed in the first article in this series (PM I; Shriberg et al., 2017a, Table 1), a highly valued attribute of diagnostic markers is generality, that is, a marker has research and clinical utility beyond identifying presence or absence of a disorder. In the present context, a behavioral marker of CAS has research and clinical generality when in addition to being diagnostically sensitive to and specific for a prior, active, and/or future disorder, it also quantifies the severity of the disorder. The three studies in the present article address that goal. We first assess the criterion validity of PM scores to index the severity

^aWaisman Center, University of Wisconsin–Madison

Correspondence to Lawrence D. Shriberg: shriberg@waisman.wisc.edu

Editor and Associate Editor: Julie Liss

Received April 12, 2016

Disclosure: The authors have declared that no competing interests existed at the time of publication.

Journal of Speech, Language, and Hearing Research • 1-17 • Copyright © 2017 American Speech-Language-Hearing Association

of CAS. We then provide findings supporting the division of PM scores into a four-classification ordinal scale of CAS severity, the PMI. We also assess the short-term and long-term stability of PM scores and the PMI in eight retrospective longitudinal case studies.

The PM

As described and referenced in the three previous articles (Shriberg et al., 2017a, 2017b, 2017c), a PM score is the percentage of occurrence of four types of inappropriate between-words pauses in a continuous speech sample. A PM score is calculated by dividing the number of such pauses in a continuous speech sample by the number of between-words pause opportunities. A PM score below 94% from a speaker of any age meets the criterion for CAS. Because inappropriate between-words pauses of the type described in PM I (Shriberg et al., 2017a) do not occur or occur only infrequently in typical speakers of either sex at any age or in speakers with SD, the PM is criterion referenced rather than normative referenced. Speakers with PM scores of 94–95.9% are classified as marginally positive for CAS when the speaker also meets the Supplemental Pause Marker Signs (SPMS) criterion for CAS described in PM I, that is, positive for at least two of the three SPMS: slow articulatory rate, inappropriate sentential stress, and transcoding errors. The present question addresses the possibility of using the distribution of PM scores below the 94% cutoff to scale the severity of CAS. Marginal PM scores were not included in these analyses because their resolution depends on participants' SPMS status.

Participant Groups

Table 1 presents a description of five study groups that included 315 speakers whose speech data were used in at least one of the three studies reported in the present article. The five groups of participants included children, adolescents, and adults whose conversational speech was sampled in collaborative studies of speech sound disorders (SSD) of known and unknown origin. All participants consented or assented using forms approved by their local institutional review board and a University of Wisconsin– Madison institutional review board. Methods to obtain, reduce, and analyze the data from the conversational speech samples were the same as those described in PM I (Shriberg et al., 2017a), with data reduction completed by the same personnel.

Group 1 included two cohorts of participants with CAS in idiopathic and neurogenetic contexts who were classified as positive for CAS on the basis of their PM scores as described in PM I (Shriberg et al., 2017a). Group 2 included participants classified as positive for CAS on the basis of the same criterion in the context of a number of complex neurodevelopmental disorders (CND) or in cohorts of children originally in treatment for idiopathic SD. Group 3 included participants classified by the second author as meeting criteria for adult-onset apraxia of speech (AAS) consequent to a stroke (apraxia of speech [AOS]) or a progressive neurological disorder (primary progressive apraxia of speech [PPAOS]). Of the 105 PM scores from these three study groups, 88 (83.8%) met the PM criterion for nonmarginal CAS, with the remaining 17 (16.2%) meeting the PM criterion for marginal CAS as defined in PM I. Participants in Group 4 (n = 8) were classified as positive or marginally positive for CAS on the basis of at least one assessment session in retrospectively assembled longitudinal data over time intervals of less than 1 month to 7.8 years. Participants in the four cohorts in Group 5 (n = 202) were classified as having SD using procedures described elsewhere (Shriberg, Austin, Lewis, McSweeny, & Wilson, 1997). None of the participants with SD in Group 5 met PM criteria for positive or marginally positive CAS. The descriptive information in Table 1 for age, sex, and percentage of consonants correct are consistent with literature findings on these sociodemographic and speech findings for speakers with speech sound disorders (SSD) of known and unknown origin.

Study I. Criterion Validity Study to Develop an Ordinal Index of CAS Severity *Rationale*

The conventional method to determine whether scores on a new measure can be used to scale performance on a domain of interest is to compute a correlation coefficient between scores on the proposed measure and scores from a criterion, diagnostic standard severity measure. In the present context, PM scores or scores derived from the PM cannot be correlated with the Mayo Clinic System criterion classification outcomes described in PM I (Shriberg et al., 2017a) because that system yields categorical classification of a speaker as meeting or not meeting criteria for CAS. In lieu of the availability of a well-validated existent intervalor ordinal-level measure of CAS severity, the criterion validity analyses in the present study were based on data from several indices and individual speech, prosody, and voice signs associated with CAS in previous studies.

As in other disorders, relatively more participants with CAS had higher PM scores (i.e., consistent with mild expression of the disorder) and relatively few had very low PM scores (i.e., consistent with severe expression of the disorder). A series of analyses of measures in our SSD databases indicated that although routinely treated as interval data in the SSD literature, most measures of speech, prosody, and voice competence, precision, and stability in participants with SSD also are not normally distributed. Two alternatives to address this psychometric need were to use arbitrary transformations to normalize each distribution of a speech, prosody, and voice measure or to use lowersensitivity Spearman association statistics rather than conventional Pearson coefficients for all correlational analyses. Both alternatives were tested and found problematic. No single distribution was appropriate to normalize each of the signs. To provide the variance needed to obtain z-scores

				Number o with P	of participan PM+ scores	ts	 4	Particip age (yea	ant ars)	%		PCC (%)		
Group	Title	Cohort	Abbreviation	Nonmarginal	Marginal	Total	М	SD	Range	males	М	SD	Range	
1	Childhood apraxia of speech (CAS) $(n = 37)$	Idiopathic CAS	CASI	21	1	22	9.1	5.1	4–23	63.6	69.8	12.1	36.8–94.3	
		Neurogenetic CAS	CASN	13	2	15	10.9	3.9	5–19	46.7	76.4	10.3	53.6-92.2	
	Total	3		34	3	37	9.8	4.7	4-23	56.8	72.5	11.8	36.8-94.3	
2	CND and PM+ Others $(n = 46)$	22q11.2 Deletion syndrome	22q11.2	3	1	4	10.3	1.7	8–12	100.0	75.7	16.1	61.1–92.4	
		Down syndrome	DS	11	4	15	14.7	1.5	12–17	40.0	79.4	7.8	59.9-89.2	
		Fragile X	FRAX	1	2	3	15.3	3.1	12–18	100.0	93.5	2.8	90.7–96.3	
		Galactosemia	GALT	5	1	6	9.3	4.8	5–16	83.3	73.9	18.7	46.5–97.8	
		Traumatic brain injury	TBI	2	0	2	3.0	0.0	NV	100.0	76.9	9.1	70.4-83.3	
		Suspected CAS	CAS-S	11	2	13	6.7	2.1	3–10	84.6	71.9	11.5	54.9–95.8	
		SD clinical cohort	SD1	1	0	1	3.0	NV	NV	100.0	89.8	NV	NV	
		SD research cohort	SD2	1	1	2	5.5	0.7	5–6	50.0	84.5	3.6	81.9-87.0	
	Total			35	11	46	5.9	2.3	3–10	71.7	74.9	11.4	54.9–95.8	
3	AAS (n = 22)	Apraxia of speech	AOS	9	1	10	63.3	12.9	45-82	70.0	89.3	7.7	68.9–95.1	
		Primary progressive AOS	PPAOS	10	2	12	70.7	10.1	53–84	50.0	91.9	6.8	74.0–97.6	
	Total			19	3	22	67.3	11.8	45–84	59.1	90.7	7.1	68.9–97.6	
4	Longitudinal participants $(n = 8)$	Clinical cohort	Longitudinal	a	a	8	15.7	15.3	6–49	42.9	78.6	10.0	66.7–89.7	
5	SD: $PM-(n = 202)$	Random cohort	SD1	0	0	0	4.4	1.3	3–9	74.1	73.2	12.6	17.5-99.1	
	,	Research cohort	SD2	0	0	0	5.5	0.6	5-7	78.9	82.6	6.4	68.8–91.3	
		Research cohort	SD3	0	0	0	4.0	0.7	3–5	73.6	70.0	9.6	36.2-87.2	
		Research cohort	SD4	0	Ō	0	4.5	0.9	3–7	48.3	68.8	11.4	42.1-82.8	
	Total			0	Ō	Ō	4.4	1.1	3–9	70.6	72.3	11.5	17.5–99.1	

Table 1. Description of five study groups that include 315 speakers whose speech data were used in at least one of the three studies.

Note. PM+ = positive for pause marker; NV = no value; SD = speech delay; PM- = negative for pause marker. PCC = percentage of consonants correct; CND = Complex neurodevelopmental disorders; AAS = Adult-onset apraxia of speech.

^aSee Table 3 for marginal or nonmarginal scores at each assessment session.

from reference data in which all scores were 0 or 100%, we added or subtracted 0.5, respectively, to one reference score, creating the marginally larger or smaller mean and the nonzero standard deviation needed for *z*-score transformations. The third alternative, selected for the present analyses, was to use Pearson correlation coefficients, which are generally robust to nonnormal distributions, with the original percentage data for the criterion validity analyses but to recommend ordinal-level analyses in research and clinical applications of the scale to index CAS severity.

Method

PM scores for 85 participants (Table 1), that is, 66 participants with CAS from Groups 1 and 2 and 19 participants from Group 3, met eligibility criteria for a criterion validity correlational analysis. Eligible scores were nonmarginal positive results for PM (PM+) to restrict the analysis to only the PM scores of interest (PM+) and to eliminate possible influence from SPMS findings.

Table 2 is a summary of Pearson correlation findings indicating the strength of associations between PM+ scores and scores on variables that assess the competence, precision, and stability of a speaker's speech, prosody, and voice. The goal of these analyses was to assess the validity of the hypothesis that PM scores have only low to moderate association with speech competence, as assessed by the 38-sign Speech Competence Index (SCI), and precision and stability as assessed by the 32-sign Precision-Stability Index (PSI; Mabie & Shriberg, 2017; Shriberg, Strand, & Mabie, 2017). Specifically, we hypothesized that only a small percentage of variance in PM scores is shared with variance in these measures of speech competence, precision, and stability, with the more significant proportion of variance independent of or orthogonal to the pathophysiological correlates of low PM scores.

The first three rows in Table 2 provide correlational findings for the PM scores with the SCI and PSI and the two measures with each other for the CAS and AAS participants. As described presently, the other 18 data rows in Table 2 were obtained from the individual SCI and PSI items and several hundred other coefficients between PM scores and individual signs of reduced speech, prosody, and voice competence, precision and stability available in the Programs to Examine Phonetic and Phonologic Evaluation Records (PEPPER; Shriberg et al., 2010) outputs. To be included in Table 2, the magnitude of the coefficient (r) for at least one of the two participant groups (CAS and AAS) had to be at least 0.32 (i.e., the PM and the sign had to share at least 10% common variance, r^2). As shown in Table 2, additional information for the SCI and PSI and the 18 signs that met the 10% or greater common variance criteria includes their linguistic domain (speech and prosody), classification analytic (competence, precision, and stability), and method of data reduction (perceptual or acoustic).

Standardization

Before proceeding to results in Table 2, it is important to underscore the effects of alternative standardization procedures on the distributions of *z*-score signs in the two indices and individual signs in Table 2. As indicated previously, the PM is a criterion-referenced rather than normative-referenced marker of CAS, that is, scores <94% meet the criterion for nonmarginal CAS without reference to a speaker's age, sex, or etiological context. The individual signs in the SCI and PSI, are standardized using *z*-scores based on a speaker's sex and chronological age.

Standardization procedures have also attempted to minimize the standard error of the mean for each z-score by increasing the age-sex sample sizes for each of the normative reference databases. The reference database of 150 typical speakers that was used to standardize speech, prosody, and voice signs of speakers 3 to 17 years of age includes five speakers of the same sex at each age (Potter et al., 2012). To minimize the standard error of the mean for this database, participants were compared with speakers of the same sex who were the same age and 1 year younger (i.e., 10 speakers total). The reference database of 50 speakers used to standardize signs for speakers ages 20 to 80 years includes 4 speakers of each sex at each decade (e.g., 20 to 29 years old) across each decade from 20 to 69 years and 5 speakers of each sex from 70 to 79 years of age (Scheer-Cohen et al., 2013).

Results

Three sets of findings in Table 2 warrant comment. The first set of findings concerns associations between PM scores and total scores on the SCI and PSI. As shown in the first three rows in Table 2, the moderate coefficients for the SCI and the PSI were significant only for the participants with CAS (r = .608; p < 0.000). None of the four coefficients assessing the association of the PM with the two measures was statistically significant at the 0.05 alpha level. These latter findings are interpreted to support the hypothesis that the inappropriate between-words pauses that comprise the PM are substantially independent of the developmental mastery of speech or prosody-voice domains as assessed in the two measures. That is, they appear to support the relative independence of a speaker's PM scores from their competence, precision, and stability as assessed by the two measures.

The second set of findings in Table 2 indicate that PM scores are also not substantially associated with any one competence, precision, or stability sign. The magnitudes of the coefficients meeting the 0.32 (i.e., at least 10% common variance) criterion for inclusion in the table ranged from 0.321 (lowered F2 on high vowels—a correlate of nasopharyngeal resonance; Fourakis, Karlsson, Tilkens, & Shriberg, 2010) to 0.597 (lengthened consonant clusters).

The third set of findings in Table 2 supports the representativeness of PM+ scores relative to their occurrence across linguistic domains, data reduction methods, and participant differences. Individual signs with the highest association with the PM come from six of the 10 linguistic domains described by Shriberg et al. (2010): vowels (three signs), consonants (seven signs), rate (two signs), and

⁴ Journal of Speech, Language, and Hearing Research • 1–17

Table 2. Pearson *r* correlation coefficients for two indices and 18 signs of speech and prosody deficits in 66 participants meeting nonmarginal pause marker (PM) criteria for childhood apraxia of speech (CAS) and 19 participants meeting nonmarginal PM criteria for adult-onset apraxia of speech (AAS).

	Compet	ence-precie	ion	м	ethod			Gro	oup	
	stabi	lity analytics	S	Perceptu	al		CAS (ı = 66)	AAS (n	e = 19)
Source	Competence	Precision	Stability	Transcription	PVSP	Acoustic	r	р	AAS (n) r .251 .380 .421 .572 .503 .567 .385 .357 .328 398 435 .597 .473 .539 .454 .419 .321	p
Indices										
Speech competence	Х			Х	Х	Х	.116	.354	.251	.300
Precision stability		Х	Х	Х	Х	х	.189	.129	.380	.109
SCI with PSI							608	000	421	073
Speech							.000	.000	. 44 1	.070
Vowels										
Lengthened diphthongs		Х				х			.572	.010
and mid vowel		X				X			500	.010
vowels		Х				Х			.503	.028
Lengthened syllables		Х				Х	.343	.005	.567	.011
Percent consonants	Х			Х					.371	.118
Percent consonants	Х			Х					.385	.104
correct—late 8 Absolute substitution	Х			Х					.357	.133
index Absolute distortion		x		x					328	170
index—late 8		X X		X		X			.020	
Square root second moment for /s/ initial singleton		Х				Х			398	.102
More diffuse /s/ initial,		Х				Х			435	.063
Lengthened consonant		Х				Х	.441	.001	.597	.007
Clusters										
Bate										
Slow	X					x	412	001		
Slow Articulation	Λ	x				X	.336	006	473	041
Stress		Λ				Λ	.000	.000	.+70	.041
Inappropriate	Х				Х				.539	.017
Percent appropriate	Х				Х				.454	.051
Jitter		Х				Х			.419	.074
Shimmer		X				X			.384	.105
Harmonics-to-noise ratio		X				X			.405	.086
Resonance quality										
Lowered F2 high vowels		Х				Х			.321	.180

stress (one sign), laryngeal quality (four signs), and resonance quality (one sign). In the Analytics and Method sections of Table 2, various signs meet the criteria for entry in Table 2 from the competence (six signs) and precision (12 signs) analytics, and at least two signs each were included using transcription (four), prosody–voice screening profile (PVSP; Shriberg, Kwiatkowski, & Rasmussen, 1990; two), and acoustic (12) methods. The number of signs meeting inclusionary criteria for the table was higher for participants in the AAS group (16 signs) than for the CAS group (four signs), with three of the 18 entries occurring for participants in both groups.

Summary

Findings in Table 2 are interpreted as criterion validity support for the PMI to scale the severity of apraxia of speech. These data indicate that as computed using Pearson coefficients, the frequency of occurrence of inappropriate between-words pauses (i.e., the PM score) is low to moderately associated with commonly reported and other signs of CAS and AAS. Thus, the PM appears to quantify a unique behavior that in PM I (Shriberg et al., 2017a) appears to be sensitive to CAS and AAS and specific for CAS in comparison to speakers with SD.

Study II. Development and Validation of the PMI

Rationale and Method

The goal of the second study was to develop and validate, for research and clinical applications, an ordinal classification of PM scores to scale the severity of CAS. Ordinal classification systems provide a way to make research or clinical comparisons between and among people with approximately similar severity of expression of a disease or disorder. In the present case, frequency of occurrence of an atypical behavior (i.e., inappropriate betweenwords pauses) is used as the basis for scaling severity of expression of apraxia of speech. A PM percentage score is the percentage of occurrence of four types of inappropriate between-words pauses in conversational or continuous speech. Frequent inappropriate cessations of speech are presumed to be the consequence of one or more deficits in speech processing, with implications for an eventual account of the functional neurobiology of these deficits. Some alternative criteria to scale severity of CAS include topographical rather than frequency of occurrence features of CAS, the number and types of additional challenges associated with CAS, and self-reported or observational instruments that classify or rate the impact of the disorder on educational, social, or vocational activities.

In summary, construct validity and psychometric support for the PM was reported in the second article in this series (PM II; Shriberg et al., 2017b), and its theoretical coherence with speech processing deficits in CAS was addressed in the third article in this series (PM III; Shriberg et al., 2017c). The goal of the present research was to assess one feature of the generality of the PM (see PM I, Table 1; Shriberg et al., 2017a): its ability to scale severity of CAS. The goal of the following analysis was to determine whether the distributional properties of PM scores meet criteria for scaling as an ordinal metric. The database used for this second study includes the PM scores from most of the same participants with CAS and AAS as described in PM III (Shriberg et al., 2017c). All participants had nonmarginal PM scores (<94.0%), meeting the criteria for apraxia of speech defined in PM I.

Results

Figure 1A is a scatterplot of the 69 PM scores from participants with CAS, with scores arranged from lowest (62%) to highest (94%). The legend indicates the group origin of each data point, including PM scores for participants with CAS-idiopathic (CASI; n = 21), CAS-neurogenetic (CASN; n = 13), Down syndrome (DS; n = 11), galactosemia

(GALT; n = 5), and other disorders (Other; n = 19). Participants in the Other group, aggregated to allow for betweengroup statistical comparisons, include participants with 22q11.2 deletion (velocardiofacial) syndrome (n = 3), traumatic brain injury (n = 2), and fragile X syndrome (n = 1) and children positive for the PM with no known etiology (n = 13). Figure 1B includes nonmarginal PM scores for 19 AAS participants, including participants with AOS (n = 9) and participants with PPAOS (n = 10).

Rationale for the division of the 69 scores in Figure 1A into the four severity categories shown was based on five characteristics of the distribution of PM scores in Figure 1A. First, the 32% point spread between the highest (94%) and lowest (62%) PM scores appeared to have sufficient sensitivity to individual differences in CAS. Second and crucial to the analysis, normality tests indicated that PM scores were not normally distributed within this range, with approximately 50% of scores ranging from 90% to 94% and having atypical variance, skew, and kurtosis values. Third, the distribution of scores as shown in Figure 1A was consistent with a three- to five-category severity scale using conventional adjectives (mild, moderate, and severe). Fourth, the "knees" of the distribution suggested that the best fit to the data would be a four-category scale at the percentile breakpoints shown across the top of each part of Figure 1A, using the corresponding suggested adjectives shown for the four classifications: mild, mild-moderate, moderate-severe, and severe. Last, the four-category severity scale, that is, the PMI, was also a good fit to the smaller distribution of PM scores from the two cohorts of 19 speakers with AAS (see Figure 1B).

Cognitive and Language Findings

Figure 2 includes intellectual and language status findings for subsets of the participants with CASI and CASN and subsets of the participants with CND in relation to their PMI classifications. Panels A and B include findings for the participants who had been administered the Kaufman Brief Intelligence Test-2 (Kaufman & Kaufman, 2004) or other standardized measures of intelligence, and Panels C and D include findings from participants who had been administered the Oral and Written Language Scales (Carrow-Woolfolk, 1995). The box plots for each severity comparison include median symbols and connecting lines, the interquartile range, and all participant data points. The general flatness of the median connecting lines in each panel suggested that participants' PMI classifications from mild to severe CAS were independent of their cognitive and language (Oral and Written Language Scales oral composite standard score) status. Parametric (analyses of variance) and nonparametric (Kruskal-Wallis) analyses of the data in each panel confirmed the visual impression that severity of CAS as classified using the PMI was not associated with intellectual or language status as assessed in this study. None of the *p* values were significant at even liberal (<.10) alpha levels appropriate for initial studies with low power. As with the previous interpretation of findings, findings in Figure 2 also are consistent with a

⁶ Journal of Speech, Language, and Hearing Research • 1–17

Figure 1. Nonmarginal pause marker (PM) scores from participants with childhood apraxia of speech (CAS; Panel A) and adult-onset apraxia of speech (AAS; Panel B) arranged from lowest to highest. AOS = apraxia of speech; CASI = CAS-idiopathic; CASN = CAS-neurogenetic; DS = Down syndrome; GALT = galactosemia; PPAOS = primary progressive AOS.



claim that a PMI classification of severity of expression of CAS is not mediated by a speaker's cognitive or language status.

PMI Findings

Figure 3 is a graphic summary of the percentages of participants in each of the four PMI classifications in three

participant groups divided into six diagnostic categories. To provide a sufficient number of participants in the CND groups, percentages are provided for the 11 participants positive for CAS in the Down syndrome group and the 11 participants positive for CAS across four other types of CND: galactosemia (five participants), 22q11.2 deletion syndrome (three), traumatic brain injury (two), and

Figure 2. Intellectual and language status findings for subsets of participants with CAS-idiopathic (CASI) and CAS-neurogenetic (CASN) childhood apraxia of speech (CAS; Panels A and C) and subsets of the participants with CAS in the context of complex neurodevelopmental disorders (Panels B and D) as classified by pause marker index (PMI) status. The box plots for each severity comparison include median symbols and their connecting lines, the interguartile range, and all participant data points.



Intellectual

fragile X syndrome (one). Two findings in Figure 3 are interpreted as criterion validity support for the PMI.

First, the percentages of participants in each of the four PMI severity classifications are broadly similar across the different etiologic contexts for apraxia of speech. The percentages of participants with mild apraxia of speech as classified by the PMI ranged from 40.0% to 46.2% across the six subgroups, mild-moderate from 10.0% to 54.5%, moderate-severe from 0.0% to 30.8%, and severe from 0.0%to 30.0%. These findings are consistent with a view of apraxia of speech that posits that severity of expression may be relatively independent of etiology. Specifically, the PMI as a severity metric may be more tied to biological correlates of inappropriate pauses than severity scales based on listener ratings or self-perceived ratings of handicap.

Second, the largest apparent difference in severity trends among the six participant subgroups occurred in the participants with Down syndrome. For these 11 participants

classified as positive for apraxia of speech, all were classified as mild or mild-moderate. As discussed by Shriberg et al. (2016), such initial trends in the severity of expression of apraxia of speech using the PMI could prove informative for research on genomic and neural substrates of CAS that crosses conventional clinical classifications (Bishop, 1997, 2015).

From a clinical perspective, inappropriate pauses that occur on at least 15% of between-words opportunities (see cutoff percentages for moderate-severe and severe PMI classifications in Figure 1) are likely perceived by both speakers and listeners as significantly disruptive to discourse, especially long-duration pauses. The PM technical report (Tilkens et al., 2016) includes tabular summaries of the durations of appropriate and inappropriate pauses blocked on relevant independent variables. As shown in the longitudinal CAS case studies and as described in the PPAOS literature (e.g., Duffy & Josephs, 2012; Duffy et al., 2015; Josephs

Figure 3. Percentage of participants in six childhood apraxia of speech (CAS), complex neurodevelopmental disorders (CND), and adult-onset apraxia of speech (AAS) cohorts whose CAS was classified as mild, mild-moderate, moderate-severe, or severe on the basis of the pause marker index (PMI). CAS_I = CAS-idiopathic; CAS_N = CAS-neurogenetic; AOS = apraxia of speech; PPAOS = primary progressive AOS.



et al., 2006), developmental and postonset changes in the severity of apraxia of speech as quantified by the PMI could be informative for clinical management.

Prosody–Voice Findings

Figures 4, 5, and 6 show averaged PVSP (Shriberg et al., 1990) findings for participants in the CAS, CND, and AAS groups, respectively. Participants in each group were subclassified by their PM scores into the four ordinal PMI classifications of apraxia of speech: mild (90.0–93.9%; Subgroup 1, open circles), mild-moderate (85.0-89.9%; Subgroup 2, open triangles), moderate-severe (80.0-84.9%; Subgroup 3, open diamonds), and severe (<80.0%; Subgroup 4, ×). Panel A in each figure includes summary numeric and graphic information for the prosody variables of phrasing, rate, and stress and the voice variables of loudness, pitch, laryngeal quality, and resonance (QUAL is a combination of the last two variables). Panel B in each figure provides percentage data on the PVSP codes used to judge the types of inappropriate prosody, which is the focus of the present analyses (i.e., voice sign findings are more relevant to the dysarthrias than to apraxia). PMI classifications of participants were based solely on their PM scores (i.e., participants with marginally positive PM scores that were resolved by meeting SPMS criteria were excluded from these analyses). Where cell sizes permitted, subgroup differences were tested using Kruskal-Wallis One-Way Analysis of Variance by Ranks (Siegel & Castellan, 1988).

Three findings across Figures 4, 5, and 6 are interpreted as criterion validity support for the PMI to scale severity **Figure 4.** Prosody–voice screening profile (PVSP) findings for participants in the childhood apraxia of speech (CAS) group. See text for description of participants and variables. Four subgroups are shown as open circles (Subgroup 1), open triangles (Subgroup 2), open diamonds (Subgroup 3), and × (Subgroup 4). APPROP and APPR = appropriate; PHRAS = phrasing; LDNSS = loudness; QUAL = quality; LARYN = laryngeal ; RESON = resonance; UTTS = utterances; Q = questionable; N = non-questionable, T = total.



of involvement in AOS. First, the average percentages of utterances scored as appropriate for rate and stress in each figure are closely aligned with participants' PMI classifications. As shown in Figure 4A, the average percentages of utterances with appropriate rate and stress for the participants with CAS paralleled the ordinal PMI subgroup classifications for these participants. Analysis of variance findings indicated that the mean percentage of utterances with appropriate rate and stress for at least one PMI subgroup differed significantly (p < .05) from the mean percentage of at least one other PMI subgroup. The alignment of appropriate rate and stress percentages with the ordinal PMI classifications was not as close for the participants in the CND (see Figure 5) and AAS (see Figure 6) groups, but the PMI severe subgroup had the lowest percentage of appropriate prosody in three of the four comparisons.

A second source of criterion support for the PMI scale is the rate and stress findings in Panels B of Figures 4,

Figure 5. Prosody–voice screening profile (PVSP) findings for participants in the complex neurodevelopmental disorders (CND) group. See text for description of participants and variables. Four subgroups are shown as open circles (Subgroup 1), open triangles (Subgroup 2), open diamonds (Subgroup 3), and × (Subgroup 4). APPROP and APPR = appropriate; PHRAS = phrasing; LDNSS = loudness; QUAL = quality; LARYN = laryngeal ; RESON = resonance; UTTS = utterances; Q = questionable; N = non-questionable, T = total.



5, and 6. As shown in the graphic sections in each figure, the high percentage of utterances meeting criteria for inappropriate rate were primarily classified as PVSP Code 9: too slow. The utterances with inappropriate stress were due primarily to utterances meeting criteria for PVSP Code 15: excessive/equal stress. The close alignment of the scores on these prosodic variables with the mean PMI classification scores is consistent with reports on the high association of rate and stress deficits with apraxia of speech (e.g., American Speech-Language-Hearing Association, 2007).

A third observation is on the laryngeal (voice) quality and resonance findings for participants in the three groups were divided into the four PMI classifications. As with the prosody variable of phrasing, participants in the three groups did not have many utterances coded as inappropriate in loudness or pitch. The CND group that was classified as severe based on the PMI had a mean of fewer than 80% of utterances coded appropriate for loudness (see Figure 5). The CND group also had more participants at all PMI severity levels with inappropriate resonance, which is consistent **Figure 6.** Prosody–voice screening profile (PVSP) findings for participants in the adult-onset apraxia (AAS) of speech group. See text for description of participants and variables. Four subgroups are shown as open circles (Subgroup 1), open triangles (Subgroup 2), open diamonds (Subgroup 3), and × (Subgroup 4). APPROP and APPR = appropriate; PHRAS = phrasing; LDNSS = loudness; QUAL = quality; LARYN = laryngeal ; RESON = resonance; UTTS = utterances; Q = questionable; N = non-questionable; T = total.



with the velopharyngeal deficits associated with many of the CND found among these participants. Even with the limitations in subgroup sizes in Figures 4, 5, and 6, the severe PMI group generally had the most utterances coded as inappropriate for laryngeal (voice) quality and resonance. Additional relevant syllable rate and pause time/syllable data for participants in the four PMI groups are provided in the PM technical report (Tilkens et al., 2016).

Summary

The PMI, which is a severity scale for CAS that divides PM scores into classifications of mild, mild-moderate, moderate-severe, and severe, is proposed here as a useful research and clinical measure. Criterion validity findings are interpreted as support for the PMI as a metric of CAS that is independent of intellectual status and language status. Additional research is needed to assess the congruence of

the four PMI classifications with alternative severity constructs (e.g., the perceived impact of CAS on communication) and to assess the temporal stability of PM scores and PMI severity classifications. The following data provide a preliminary estimate of the latter psychometric property.

Study III. PMI Findings from Eight Retrospective Longitudinal Case Studies *Rationale*

The goals associated with the PM and the PMI were to provide a single-sign marker of CAS for research leading to a biomarker of CAS and (as discussed in PM III; Shriberg et al., 2017c) for treatment applications. An estimate of the test-retest stability of PM scores and the four proposed PMI severity classifications will require a sufficiently large sample of test-retest scores representative of persons with CAS obtained within a short test-retest interval (i.e., to rule out changes due to normalization with or without treatment). Retest reliability data can then be used to estimate the standard error of measurement of the PM percentage scores. Until such data are available, it is useful to provide a preliminary estimate of the test-retest stability of PM scores and the PMI from a small sample of retrospective case studies.

Method

Audio-recorded conversational speech samples from collaborative research and treatment sessions at two university speech clinics were used to obtain samples eligible for inclusion as retrospective longitudinal case studies. The inclusionary criteria were (a) speakers of any age positive for CAS as determined by the PM on at least one of a minimum of two sessions and (b) the audio samples from eligible participants met technical quality and linguistic content criteria for analyses using the data reduction procedures for transcription, prosody–voice coding, and acoustic analyses described in PM I (Shriberg et al., 2017a).

Table 3 includes a description of findings for eight participants whose total of 23 audio samples met the two inclusionary criteria. Two to seven audio samples were available for each of the participants. Original analog recordings for several of the participants were digitized for the present analyses using standard methods. Using similar procedures, the authors who reduced the data for PM I (Shriberg et al., 2017a) and analyses in the present article reduced these data and obtained PM scores for each of the 23 audio files.

Participants' rounded ages at the first recording session ranged from 6 to 49 years. Five of the eight participants (62.5%) were female. The six youngest participants had received or were receiving treatment for CAS or suspected CAS, with less than 4 weeks to 7.5 years between the samples from each of the eight participants. Examiners who obtained the speech samples were either student clinicians or their clinical supervisors. The lengths of the samples differed considerably, ranging from 1 to 6 min. Most samples were speech probes used to track treatment progress, and they tended to be shorter than a typical sample obtained in an initial clinical assessment session. The two assessment sessions for each of the two oldest participants, obtained within a 5-month period, were conducted at two sites as part of a research study in CAS. Each participant was assessed by a different examiner at each site.

Table 3 also includes findings for the three SPMS obtained from the speech samples. Transcoding findings, obtained from responses to the Syllable Repetition Task (Shriberg et al., 2009), were available for only one sample from each of six participants. Table 3 also includes data for each session, the PM score, the PM-SPMS and PM classification outcomes, and the CAS classification and severity adjective classification outcome, which were based on the PM and the PMI, respectively.

Results

Two observations on the findings for each of the eight participants in Table 3 are interpreted as support for the short- and long-term stability of PM scores and PMI classifications.

Entries in the "Pause Marker Classification" and "CAS Classification" columns generally support the stability of session-to-session PM scores and the PMI. Although the data are minimal for inferential statistical testing, PM scores and especially the PMI classifications appear to be generally similar for participants across time. Where sessionto-session differences occur, they were generally in the expected direction toward milder involvement, with positive change in this uncontrolled study best interpreted as likely due to growth and development and to treatment efficacy. Negative PM and PMI changes (e.g., Participants 4 and 7) could accurately reflect participant status or could indicate measurement error. The seven PM scores and PMI classifications obtained for Participant 3, whose CAS is associated with a genetic polymorphism, provide strong support for the stability of PM scores and PMI adjectives. In the seven treatment sessions (spanning almost 8 years from 6 to approximately 14 years of age), his PM scores and PMI classifications indicate progression from severe to mild CAS.

A second observation from the findings in Table 3 concerns the validity of the SPMS procedure to resolve marginal PM scores and more generally to support all CAS classification outcomes. Notwithstanding the missing data on transcoding outcomes, participant status on slow articulatory rate and inappropriate sentential stress were generally concordant with participant PM scores, thus supporting the validity of both the marginal and nonmarginal CAS scores and their corresponding severity classifications.

Lacking the controls designed for prospective studies and the sample sizes required for inferential statistics, the retrospective longitudinal case study findings for the PM scores and PMI classifications in Table 3 are interpreted as at least proof-of-concept support for the temporal stability of the PMI.

	Age (vears:			Supplementary PM signs (SPMS) classification								
		Age		Slow	Inappropriate			PM			CAS	
Participant	months)	(years)	Sex	rate	stress	Transcoding	SPMS	% score	Description	Classification	Description	Severity
1	5;6	6	F	+	+	ND	SPMS+	100.0	PM-SPMS+	PM-	CAS-	-
1	5;7	6	F	+	-	ND	SPMSI	97.8	PM-SPMSI	PM-	CAS-	_
1 ^a	6;4	6	F	+	+	+	SPMS+	(95.5) ^b	(PM)SPMS+	(PM+)	(CAS+)	Mild
2	5;9	6	F	_	_	ND	SPMS-	91.8	PM+SPMS-	PM+	CAS+	Mild
2 ^a	6;11	7	F	-	+	+	SPMS+	98.2	PM-SPMS+	PM-	CAS-	-
3	6;0	6	М	+	+	ND	SPMS+	78.0	PM+SPMS+	PM+	CAS+	Severe
3	6;0	6	Μ	+	+	ND	SPMS+	79.3	PM+SPMS+	PM+	CAS+	Severe
3 ^a	11;1	11	Μ	+	+	-	SPMS+	62.1	PM+SPMS+	PM+	CAS+	Severe
3	11;2	11	Μ	+	+	ND	SPMS+	84.6	PM+SPMS+	PM+	CAS+	Moderate-severe
3	11;4	11	Μ	+	+	ND	SPMS+	(94.4) ^b	(PM)SPMS+	(PM+)	(CAS+)	Mild
3	13;8	14	Μ	+	+	ND	SPMS+	(94.4) ^b	(PM)SPMS+	(PM+)	(CAS+)	Mild
3	13;9	14	Μ	+	+	ND	SPMS+	93.3	PM+SPMS+	PM+	CAS+	Mild
4	7;9	8	М	_	_	ND	SPMS-	82.0	PM+SPMS-	PM+	CAS+	Moderate-severe
4	7;9	8	М	+	+	ND	SPMS+	79.5	PM+SPMS+	PM+	CAS+	Severe
5 ^a	10;11	11	М	_	+	+	SPMS+	93.8	PM+SPMS+	PM+	CAS+	Mild
5	12;0	12	М	+	-	ND	SPMSI	93.3	PM+SPMSI	PM+	CAS+	Mild
6 ^a	12;3	12	F	+	+	+	SPMS+	90.3	PM+SPMS+	PM+	CAS+	Mild
6	13;5	13	F	+	_	ND	SPMSI	93.4	PM+SPMSI	PM+	CAS+	Mild
6	13;5	13	F	+	-	ND	SPMSI	100.0	PM-SPMSI	PM-	CAS-	-
7	18;5	18	F	+	+	_	SPMS+	(95.7) ^b	(PM)SPMS+	(PM+)	(CAS+)	Mild
7 ^a	18;10	19	F	+	+	+	SPMS+	89.6	PM+SPMS+	PM+	CAS+	Mild-moderate
8	49;4	49	F	+	+	_	SPMS+	(94.5) ^b	(PM)SPMS+	(PM+)	(CAS+)	Mild
8	49;9	50	F	+	+	+	SPMS+	92.4	PM+SPMS+	PM+	CAS+	Mild

Table 3. Description of findings for the eight participants with longitudinal pause marker (PM) data.

Note. F = female; M = male; ND = no data; SPMSI = SPMS indeterminate.

^aPM% score obtained for the present study. ^bMarginal.

Conclusion

Findings reported in PM II, PM III (Shriberg et al., 2017b, 2017c), and the present article are interpreted to be consistent with six of the seven attributes of valued diagnostic markers described in PM I (Table 1; Shriberg et al., 2017a): accuracy, reliability, coherence, discreteness, parsimony, and generality. The psychometric findings and findings discussed below that presently constrain the external validity of the PM warrant classification of the PM as a "near conclusive" rather than conclusive diagnostic marker to discriminate CAS from SD. Support for the seventh attribute, efficiency, will require additional study to automate data collection and data reduction.

The final section of this series addresses PM research needs and directions that have arisen across the four articles in this series. The discussion includes clearly speculative comments and suggestions.

Research Directions

Findings presented in this four-part series are interpreted as support for viewing CAS as a multidomain disorder with core deficits in both the speech processes required to acquire, store, and retrieve phonological representations and the processes that plan, program, and monitor articulatory gestures and their products. Findings were interpreted as support for the theoretical coherence of the PM and PMI with the two core processing deficits in CAS. We suggest that the PM provides a behavioral marker of apraxia in which inappropriate pauses and the speech gestures that may occur before, during, and after are the behavioral signs of representational and transcoding deficits, respectively.

PM Research Needs

An important constraint on the PM findings reported in these articles are the lack of data from participants with the most severe expression of CAS, effectively truncating the lower boundaries of the distributions of relevant speech sign variables and the distribution of PM scores. Effective procedures are needed to obtain linguistically adequate continuous speech samples from children with low verbal skills who may produce some phrase-length speech with supportive assessment. Repeated continuous speech sampling in environmentally valid contexts using mobile recording instrumentation is becoming an option, with the goal of providing automated data acquisition and reduction for the types of speech analyses described in the present article (e.g., Dykstra et al., 2013; Shahin et al., 2015). It is also possible that the pause and speech elements of the PM are occurring in other simpler speech tasks in the present assessment protocol or could be evoked using other speech tasks that do not require continuous speech (Kearney et al., 2015). In all contexts, signal processing software is needed to identify and quantify both the occurrences and durations of inappropriate pauses in an audio file and the abrupt speech onsets (and offsets) of phonemes in abrupt inappropriate pauses.

An additional constraint on the present findings is the lack of subgroup data within each of the three groups of participants with apraxia of speech. To increase cell sizes, the CAS group included participants with CASI and CASN, the CND group included participants with CAS in the context of a range of neurodevelopmental disorders, and the AAS group included participants with AOS and PPAOS. Close analyses of the nonaggregated findings for participants in these subgroups will likely inform questions about core processing deficits in CAS and could lead to improved sensitivity and specificity for the PM. As mentioned in PM II, the insensitivity of the PM for a "voicers" subgroup of participants with AAS needs to be addressed for its possible theoretical informativeness, with implications for treatment.

As discussed in several places in this series of articles, constraints associated with the behavioral methods and descriptive research design of the CAS studies have limited both the questions posed and the grain size of the findings reported. A major need is to assess the specificity of the PM with other samples of speakers with SSD, including speakers across developmental epochs with representative subtypes and severity levels of dysarthria, stuttering, cluttering, and velopharyngeal incompetence. Findings do not include information on somatosensory representational processes (as shown in Figure 1), on planning versus programming stages of transcoding deficits, or on feedforward or feedback deficits in CAS. PM research using neuroimaging and electrophysiological modalities and other instrumental methods is needed to cross-validate, extend, and explicate the present findings and claims.

PM Research Questions

The signature sign of apraxia in AAS is a unique type of inconsistent, effortful production of challenging multisyllabic words (Ziegler & Aichert, 2015). There are several problems with attempting to evoke multisyllabic words from children with CAS, primarily associated with limitations in language and phonetic inventory. The pause and speech elements that make up the PM are proposed as potential signatures of CAS that are independent of age, sex, etiology, cognitive status, language status, and severity of speech involvement. One research question to address is whether PM and PMI classifications also are independent of a speaker's first or subsequent language and the dialect of the language. Following are some comments and research ideas for the eight inappropriate pause types in the PM.

Inappropriate Pauses

A primary goal in CAS research, as in research for all disorders and diseases, is to identify a biomarker for early detection, prediction, mitigation, and possible prevention. We posit that similar to a stutter in dysfluency, the pause element of the PM is sensitive to the moment

of apraxia—the point in talking when preexecution commands are not sufficient to continue speaking. The present findings are viewed as support for the potential of research on this moment to explicate the neurocognitive and neuromotor correlates of CAS across the life span.

Abrupt Inappropriate Word Onsets/Offsets

The second element of the PM, the speech events before, during, or after the pause, provides the basis for the eight inappropriate pauses typology. We speculate that the brief speech events in the most frequent type of inappropriate pauses, that is, abrupt inappropriate pauses, could provide the biological event needed for a biomarker of CAS in the context of signal processing technology. Momentary (within a few milliseconds) increases in amplitude of the inappropriate between-words abrupt pauses are consistent with deficits associated with state feedback control theory (Houde & Nagarajan, 2011), with possible control parameters including movement commands for respiratory, laryngeal, and articulatory targets. Phonatory behaviors as biomarkers of apraxia of speech, in particular, have attractive measurement features as they do in research in voice, dysfluency, and other motor speech disorders (e.g., Civier, Bullock, Max, & Guenther, 2013; Cohen, Renshaw, Mitchell, & Kim, 2016; Kim, 2015; Konopka & Roberts, 2016; Kumar, Croxson, & Simonyan, 2016; Ludlow, 2015; Neef, Anwander, & Friederici, 2015; Pouplier, Marin, & Waltl, 2014; Simonyan, 2013; Simonyan & Horwitz, 2011; Vanhoutte et al., 2014). In the present context, the same control mechanism underlying an abrupt inappropriate speech onset could underlie the excessive/equal sentential stress sign of CAS described previously. A few inappropriate abrupt pauses were also perceptually identified in a small number of participants with SD. Research is needed to determine whether abrupt pauses are similar in individuals with CAS and SD or whether there may be typological differences, with implications for theory and clinical practice.

Short-term and long-term instrumental studies of abrupt speech onsets and offsets from participants in treatment for CAS and AAS also could be informative (Maassen, 2002; Odell & Shriberg, 2001; Poole, Gallagher, Janosky, & Qualls, 1997). Significant differences in the proportion of abrupt Type I pauses were found for participants with CAS (approximately 70%) compared with participants with AAS (approximately 49%). Fine-grained analyses of the correlates of such differences might lead to predictive and prognostic information on normalization expectancies in individuals with CAS and on correlates of increasing severity in individuals with PPAOS.

Other Inappropriate Between-Words Pauses

Definitions for the eight inappropriate pause types in the PM were given in PM I (Shriberg et al., 2017a). Preliminary analyses indicated that the total percentage of occurrence of each of the four Type I inappropriate betweenwords pauses (abrupt, alone, change, and grope) was the most sensitive and specific sign of CAS relative to SD. The other four types of inappropriate pauses (Type II pauses: long, breath, repetitions/revisions, and additions) were retained for their potential to inform explanatory accounts of speech processing in apraxia of speech. In the present article, in addition to findings for abrupt inappropriate pauses in the Results section, the frequencies of occurrence of grope, repetitions/revisions, and additions were used to infer deficits in representational and transcoding processes in individuals with CAS and in discussions of feedforward and feedback processes.

One potentially informative avenue of study of all eight inappropriate pauses would be to search for commonalities among them (i.e., factor structures) that may provide insights to connectivity networks underlying segmental and suprasegmental profiles. For example, inappropriate change pauses are defined as pauses immediately preceded or followed by a phoneme or word that includes a significant change in amplitude, frequency, or rate. Preliminary analyses indicate that the most common change is in amplitude, which might invoke the same changes in force regulation (respiratory, laryngeal, or supralaryngeal) as posited to underlie inappropriate abrupt pauses.

Another potentially informative analysis for the speech processing questions discussed above is to examine the individual profiles of participants in the three groups to identify those with possible departures from the groupwise frequencies of occurrence of the eight inappropriate pause types. For example, participants with a disproportionate number of inappropriate long pauses (>750 ms) may have other characteristics in common that yield insights into neural substrates of representational and transcoding deficits in individuals with CAS. For example, a participant with significantly more frequent breath pauses (defined in PM I (Shriberg et al., 2017a) as a pause that includes audible inhalation not associated with excessive length of the utterance or emotional excitement) may have some respiratory or laryngeal signs elsewhere in the assessment data that could be of interest for explicating pathophysiological substrates of transcoding deficits.

A third possibly informative use of the frequency of occurrence of all eight subtypes of inappropriate pauses is their potential association with speech processing changes over the time course from onset to resolution of CAS. If the inappropriate pauses of grope and repetitions/revisions are posited to occur only when phonological representations are sufficiently specified, the presence and/or frequency of such inappropriate pauses could presage normalization. Consistent with one of the proposed criteria for a core process—that it persists until normalization of CAS—the most reliably persistent of the eight inappropriate pauses may provide an avenue for backward engineering of the functional biology of CAS.

PM and the PMI in Treatment of CAS

Treatment Goals

If, as posited, CAS is best characterized as a multidomain speech disorder, it follows that CAS treatment goals

should not be limited to improving speech production as a motor skill. Rather, for speakers of any age with CAS, treatment goals would more appropriately focus on deficits in subdomains of both representational and transcoding processes. Findings from a comprehensive assessment protocol that yields information on both phonological competence and motor speech skills should suggest the appropriate weighting of representational goals (i.e., to instantiate accurate linguistic representations and timely access to them) and transcoding goals (i.e., to develop the plans and programs for precise and stable speech and prosody and timely access to them). Bridging these representational and transcoding goals, a third treatment domain might address possible deficits in self-monitoring (feedforward and feedback) signals that link emerging accurate representations to emerging accurate and stable articulation and prosody. The challenge in CAS treatment with a motor speech focus is that the clinician's success in evoking and shaping increasingly correct, precise, and stable speech and prosody (i.e., shaping) is dependent on the integrity of processes that both precede (representational and feedforward) and follow (auditory and somatosensory feedback) the person's speech productions. The claim here, again, is that the present findings indicate that although motor speech goals are necessary in CAS treatment, some research supports the perspective that motor speech goals likely are not sufficient (Preston, Maas, Whittle, Leece, & McCabe, 2016; Rvachew & Brosseau-Lapré, 2012).

Treatment Targets

The eight types of inappropriate between-words pauses in the PM may be informative for selecting and prioritizing individual treatment targets. If these and possibly other types of inappropriate pauses are sensitive to the moment of apraxia, determining how often each type occurs and in what linguistic contexts could yield useful insights for treatment planning. The present research did not include individual analyses of the several hundreds of transcripts to determine whether the absolute and relative percentages of occurrence of the eight types of inappropriate pauses were associated with lexical, syntactic, or pragmatic variables or other language variables. A disproportionate percentage of one or more types of inappropriate pauses (e.g., repetitions/revisions) in a person with CAS could suggest the locus of a theoretically informative speech processing deficit and, as with primary and accessory behaviors in stuttering, could be a treatment target because of its undue negative effects on speech intelligibility or acceptability.

A second potential use of the PM for target selection follows from the present and other findings indicating that slow rate and inappropriate sentential stress are corollary deficits in individuals with CAS. In a previous report (Shriberg, Strand, & Jakielski, 2012), deficits in phrasing

encoding deficits are common to all verbal trait disorders (e.g., language impairment, SD, and reading disorder), and slow rate and inappropriate sentential stress are observed in individuals with many to most types of motor speech disorders. In the present context, research reported in PM I suggests that the single-sign PM better meets each of the seven proposed criteria for highly valued diagnostic markers proposed in that article. To resolve marginal PM scores and to strengthen the validity of PM classification decisions, slow articulatory rate, inappropriate sentential stress, and an encoding deficit as quantified by the Transcoding scale of the Syllable Repetition Task were utilized as supplemental information for PM classification (i.e., SPMS). Although not sufficiently sensitive to or specific for CAS, a speaker's quantitative status on phrasing (i.e., inappropriate pauses), rate, and sentential stress could guide target selection in the treatment of CAS in children and youth and particularly in adults with persistent CAS.

Acknowledgments

This research was supported by National Institute on Deafness and Other Communication Disorders Grant DC000496 and by a core grant to the Waisman Center from the National Institute of Child Health and Development (Grant HD03352). Authors Christie Tilkens and Heather Karlsson made the original observations and substantial continuing contributions leading to the development of the PM. We are grateful to the following colleagues and collaborators for their significant contributions to this research: Len Abbeduto, Nancy Alarcon, Becky Baas, Adriane Baylis, Richard Boada, Roger Brown, Stephen Camarata, Thomas Campbell, Richard Folsom, Lisa Freebairn, Jordan Green, Barbara Lewis, Christopher Moore, Katherine Odell, Bruce Pennington, Nancy Potter, Jonathan Preston, Erin Redle, Heather Leavy Rusiewicz, Alison Scheer-Cohen, Kristie Spencer, Ruth Stoeckel, Bruce Tomblin, Jennifer Vannest, and Emily White. We also thank the many participants, parents of participants, and research colleagues who have contributed significant insights into the needs and issues of diagnostic research in CAS.

References

- American Speech-Language-Hearing Association. (2007). Childhood apraxia of speech [Technical report]. Retrieved from http://www.asha.org/policy/TR2007-00278/
- Bishop, D. V. M. (1997). Cognitive neuropsychology and developmental disorders: Uncomfortable bedfellows. The Quarterly Journal of Experimental Psychology, 50A, 899–923.
- Bishop, D. V. M. (2015). The interface between genetics and psychology: Lessons from developmental dyslexia. Proceedings of the Royal Society B, 282(1806). https://doi.org/10.1098/rspb. 2014.3139
- Carrow-Woolfolk, E. (1995). Oral and Written Language Scales (OWLS). Bloomington, MN: Pearson Assessment.
- Civier, O., Bullock, D., Max, L., & Guenther, F. H. (2013). Computational modeling of stuttering caused by impairments in a basal ganglia thalamo-cortical circuit involved in syllable selection and initiation. Brain & Language, 126, 263-278.
- Cohen, A. S., Renshaw, T. L., Mitchell, K. R., & Kim, Y. (2016). A psychometric investigation of "macroscopic" speech measures for clinical and psychological science. Behavior Research

Duffy, J. R., & Josephs, K. A. (2012). The diagnosis and understanding of apraxia of speech: Why including neurodegenerative etiologies may be important. *Journal of Speech, Language,* and Hearing Research, 55, S1518–S1522.

Duffy, J. R., Strand, E. A., Clark, H., Machulda, M., Whitwell, J. L., & Josephs, K. A. (2015). Primary progressive apraxia of speech: Clinical features and acoustic and neurologic correlates. *American Journal of Speech-Language Pathology*, 24, 88–100.

Dykstra, J. R., Sabatos-Deviot, M. G., Irvin, D. W., Boyd, B. A., Hume, K. A., & Odom, S. L. (2013). Using the Language Environment Analysis (LENA) system in preschool classrooms with children with autism spectrum disorders. *Autism*, 17, 582–594.

Fourakis, M., Karlsson, H., Tilkens, C., & Shriberg, L. D. (2010). Acoustic correlates of nasal and nasopharyngeal resonance [Technical report 15]. Phonology Project, Waisman Center, University of Wisconsin–Madison.

Houde, J. F., & Nagarajan, S. S. (2011). Speech production as state feedback control. *Frontiers in Human Neuroscience*, 5. https://doi.org/10.3389/fnhum.2011.00082

Josephs, K. A., Duffy, J. R., Strand, E. A., Whitwell, J. L., Layton, K. F., Parisi, J. E., ... Petersen, R. C. (2006). Clinicopathological and imaging correlates of progressive aphasia and apraxia of speech. *Brain*, 129, 1385–1398.

Kaufman, A. S., & Kaufman, N. L. (2004). Kaufman Brief Intelligence Test (KBIT-2) (2nd ed.). London: Pearson.

Kearney, E., Granata, F., Yunusova, Y., van Lieshout, P., Hayden, D., & Namasivayam, A. (2015). Outcome measures in developmental speech sound disorders with a motor basis. *Current Developmental Disorders Report*, 2, 253–272.

Kim, S. I. (2015). The association between the supraglottic activity and glottal stops at the sentence level (Unpublished master's thesis). University of Iowa, Iowa City. Retrieved from http:// ir.uiowa.edu/etd/1660

Konopka, G., & Roberts, T. F. (2016). Insights into the neural and genetic basis of vocal communication. *Cell*, 164, 1269–1276.

Kumar, V., Croxson, P. L., & Simonyan, K. (2016). Structural organization of the laryngeal motor cortical network and its implication for evolution of speech production. *The Journal of Neuroscience*, 36, 4170–4181.

Ludlow, C. L. (2015). Central nervous system control of voice and swallowing. *Journal of Clinical Neurophysiology*, 32, 294–303.

Maassen, B. (2002). Issues contrasting adult acquired versus developmental apraxia of speech. *Seminars in Speech and Language*, 23, 257–266.

Mabie, H. L., & Shriberg, L. D. (2017). Speech and motor speech measures and reference data for the Speech Disorders Classification System (SDCS). (Technical Report No. 23). Retrieved from Phonology Project website: http://www.waisman.wisc. edu/phonology/.

Neef, N. E., Anwander, A., & Friederici, A. D. (2015). The neurobiological grounding of persistent stuttering: From structure to function. *Current Neurology and Neuroscience Reports*, 15, 63.

Odell, K. H., & Shriberg, L. D. (2001). Prosody-voice characteristics of children and adults with apraxia of speech. *Clinical Linguistics and Phonetics*, 15, 275–307.

Poole, J. L., Gallagher, J., Janosky, J., & Qualls, C. (1997). The mechanisms for adult-onset apraxia and developmental dyspraxia: An examination and comparison of error patterns. *American Journal of Occupational Therapy*, 51, 339–346. Potter, N. L., Hall, S., Karlsson, H. B., Fourakis, M., Lohmeier, H. L., McSweeny, J. L., ... Shriberg, L. D. (2012). Reference data for the Madison Speech Assessment Protocol (MSAP): A database of 150 participants 3-to-18 years of age with typical speech [Technical report 18]. Waisman Center, University of Wisconsin–Madison.

Pouplier, M., Marin, S., & Waltl, S. (2014). Voice onset time in consonant cluster errors: Can phonetic accommodation differentiate cognitive from motor errors? *Journal of Speech, Language, and Hearing Research, 57*, 1577–1588.

Preston, J. L., Maas, E., Whittle, J., Leece, M. C., & McCabe, P. (2016). Limited acquisition and generalisation of rhotics with ultrasound and visual feedback in childhood apraxia. *Clinical Linguistics and Phonetics*, 30, 363–381.

Rvachew, S., & Brosseau-Lapré, F. (2012). Developmental phonological disorders: Foundations of clinical practice. San Diego, CA: Plural Publishing.

Scheer-Cohen, A. R., Holt, A. S., Karlsson, H. B., Mabie, H. L., McSweeny, J. L., Tilkens, C. M., & Shriberg, L. D. (2013). Reference data for the Madison Speech Assessment Protocol (MSAP): A database of fifty 20-to-80 year old participants with typical speech [Technical report 20]. Waisman Center, University of Wisconsin–Madison.

Shahin, M., Ahmed, B., Parnandi, A., Karappa, V., McKechnie, J., Ballard, K. J., & Gutierrez-Osuna, R. (2015). Tabby talks: An automated tool for the assessment of childhood apraxia of speech. *Speech Communication*, 70, 49–64.

Shriberg, L. D., Austin, D., Lewis, B. A., McSweeny, J. L., & Wilson, D. L. (1997). The Speech Disorders Classification System (SDCS): Extensions and lifespan reference data. *Journal of Speech, Language, and Hearing Research*, 40, 723–740.

Shriberg, L. D., Fourakis, M., Hall, S., Karlsson, H. B., Lohmeier, H. L., McSweeny, J., ... Wilson, D. L. (2010). Extensions to the Speech Disorders Classification System (SDCS). *Clinical Linguistics & Phonetics, 24, 795–824.*

Shriberg, L. D., Kwiatkowski, J., & Rasmussen, C. (1990). *The* prosody–voice screening profile. Tucson, AZ: Communication Skill Builders.

Shriberg, L. D., Lohmeier, H. L., Campbell, T. F., Dollaghan, C. A., Green, J. R., & Moore, C. A. (2009). A nonword repetition task for speakers with misarticulations: The Syllable Repetition Task (SRT). *Journal of Speech, Language, and Hearing Research, 52*, 1189–1212.

Shriberg, L. D., Strand, E. A., Fourakis, M., Jakielski, K. J., Hall, S. D., Karlsson, H. B., ... Wilson, D. L. (2017a). A diagnostic marker to discriminate childhood apraxia of speech from speech delay: I. Development and description of the pause marker (PM). *Journal of Speech, Language, and Hearing Research*. Advance online publication. https:// doi.org/10.1044/2016_JSLHR-S-15-0296

Shriberg, L. D., Strand, E. A., Fourakis, M., Jakielski, K. J., Hall, S. D., Karlsson, H. B., ... Wilson, D. L. (2017b). A diagnostic marker to discriminate childhood apraxia of speech from speech delay: II. Validity studies of the pause marker (PM). *Journal of Speech, Language, and Hearing Research*. Advance online publication. https://doi.org/10.1044/2016_JSLHR-S-15-0297

Shriberg, L. D., Strand, E. A., Fourakis, M., Jakielski, K. J., Hall, S. D., Karlsson, H. B., ... Wilson, D. L. (2017c). A diagnostic marker to discriminate childhood apraxia of speech from speech delay: III. Theoretical coherence of the PM with speech processing deficits in CAS. *Journal of*

16 Journal of Speech, Language, and Hearing Research • 1–17

Speech, Language, and Hearing Research. Advance online publication. https://doi.org/10.1044/2016_JSLHR-S-15-0298

- Shriberg, L. D., Strand, E. A., & Jakielski, K. J. (March, 2012). Diagnostic Signs of Childhood Apraxia of Speech in Idiopathic, Neurogenetic, and Complex Neurodevelopmental Contexts. Paper presented at the Sixteenth Biennial Conference on Motor Speech: Motor Speech Disorders & Speech Motor Control, Santa Rosa, CA.
- Shriberg, L. D., Strand, E. A., & Mabie, H. L. (2017). Prevalence of speech and motor speech disorders in idiopathic speech delay and complex neurodevelopmental disorders. Manuscript submitted for publication.
- Siegel, S., & Castellan, N. J. (1988). Nonparametric statistics for the behavioral sciences (2nd Ed.). New York, NY: McGraw-Hill.
- Simonyan, K. (2013). The laryngeal motor cortex: Its organization and connectivity. *Current Opinion in Neurobiology*, 28, 15–21.

- Simonyan, K., & Horwitz, B. (2011). Laryngeal motor cortex and control of speech in humans. *Neuroscientist*, 17, 197–208.
- Tilkens, C. M., Karlsson, H. B., Fourakis, M., Hall, S. D., Mabie, H. L., McSweeny, J. L., ... Shriberg, L. D. (2017). A diagnostic marker to discriminate childhood apraxia of speech (CAS) from Speech Delay (SD). (Technical Report No. 22). Retrieved from Phonology Project website: http://www.waisman. wisc.edu/phonology/.
- Vanhoutte, S., Van Borsel, J., Cosyns, M., Batens, K., van Mierlo, P., Hemelsoet, D., ... Corthals, P. (2014). CNV amplitude as a neural correlate for stuttering frequency: A case report of acquired stuttering. *Neuropsychologia*, 64C, 349–359.
- Ziegler, W., & Aichert, I. (2015). How much is a word? Predicting ease of articulation planning from apraxic speech error patterns. *Cortex*, 69, 24–39.