JSLHR

Research Article

Data-Driven Subclassification of Speech Sound Disorders in Preschool Children

Jennell C. Vick,[a] Thomas F. Campbell,[b] Lawrence D. Shriberg,[c] Jordan R. Green,[d] Klaus Truemper,[b] Heather Leavy Rusiewicz,[e] and Christopher A. Moore[f]

Purpose: The purpose of the study was to determine whether distinct subgroups of preschool children with speech sound disorders (SSD) could be identified using a subgroup discovery algorithm (SUBgroup discovery via Alternate Random Processes, or SUBARP). Of specific interest was finding evidence of a subgroup of SSD exhibiting performance consistent with atypical speech motor control.
Method: Ninety-seven preschool children with SSD completed speech and nonspeech tasks. Fifty-three kinematic, acoustic, and behavioral measures from these tasks were input to SUBARP.
Results: Two distinct subgroups were identified from the larger sample. The 1st subgroup (76%; population prevalence estimate = 67.8%–84.8%) did not have characteristics that would suggest atypical speech motor control. The 2nd subgroup (10.3%; population prevalence estimate = 4.3%–16.5%) exhibited significantly higher variability in measures of articulatory kinematics and poor ability to imitate iambic lexical stress, suggesting atypical speech motor control. Both subgroups were consistent with classes of SSD in the Speech Disorders Classification System (SDCS; Shriberg et al., 2010a).
Conclusion: Characteristics of children in the larger subgroup were consistent with the proportionally large SDCS class termed *speech delay;* characteristics of children in the smaller subgroup were consistent with the SDCS subtype termed *motor speech disorder—not otherwise specified.* The authors identified candidate measures to identify children in each of these groups.

The purpose of this study was to determine whether a subgroup of children with atypical speech motor control could be identified from a sample of children with speech sound disorders (SSD). We analyzed 53 measures, including measures of speech movement, from a relatively large number of cases (N = 97). We used a subgroup discovery algorithm to achieve this aim, a technique within the data-driven methods of machine learning.

Subgroup discovery seeks to identify subgroups within a set of data without any a priori assumption of the number or size of the subgroups to be identified. A subgroup discovery method generally looks for important patterns in the data, derives rules from the patterns, and then uses the rules to characterize subgroups. The rules may take on various forms that may or may not allow human interpretation, though an important objective of this investigation was that the emergent rules could be easily understood and subjected to expert interpretation and knowledge of the selected subgroups.

Many subgroup discovery techniques are available for data mining (Herrera, Carmona, González, & del Jesus, 2011). There is a general methodology to convert machine learning techniques that explain differences between two sets into subgroup discovery techniques (Lavrač, Cestnik, Gamberger, & Flach, 2004). We selected a subgroup discovery method that was based on this construction and that would produce humanly comprehensible rules (Truemper, 2009). The method, called SUBgroup discovery via Alternate Random Processes, or SUBARP, was applied to our data from preschool children with SSD, and the results are the focus of this article.

SSD, and specifically speech delay (SD), are highly prevalent in preschool children (15.6% among 3-year-olds; Campbell et al., 2003), with approximately 4% of all children having persistent SD at age 6 years (Shriberg, Tomblin, & McSweeny, 1999). As a population, children with SSD are heterogeneous, and the presumption of distinct subgroups

[a]Case Western Reserve University, Cleveland, OH
[b]University of Texas at Dallas, Richardson
[c]Waisman Center, University of Wisconsin—Madison
[d]MGH Institute of Health Professions, Boston, MA
[e]Duquesne University, Pittsburgh, PA
[f]U.S. Department of Veterans Affairs, Washington, DC
Correspondence to Jennell C. Vick: jennell@case.edu

Editor: Jody Kreiman
Associate Editor: Ben A. M. Maassen
Received June 21, 2012
Revision received May 1, 2013
Accepted June 27, 2014
DOI: 10.1044/2014_JSLHR-S-12-0193

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

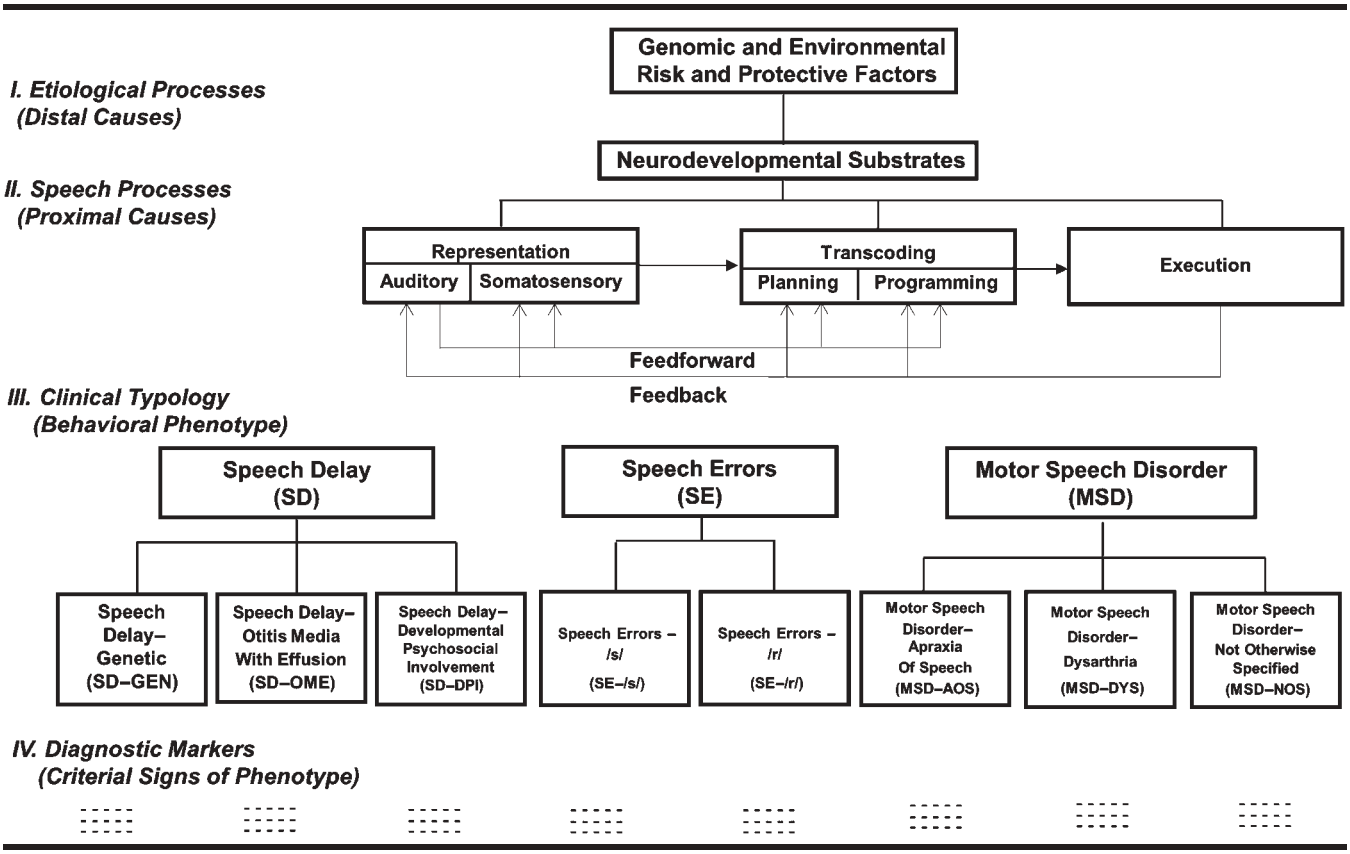
Downloaded From: http://jslhr.pubs.asha.org/ by a Health Sci Learning Ctr User on 04/14/2015
Terms of Use: http://pubs.asha.org/ss/Rights_and_Permissions.aspx

is common. Researchers have sought to identify distinct subgroups of SSD to improve prognostic accuracy and to motivate more narrowly targeted interventions. Historically, children with SSD were classified under the category of phonetic-based articulatory disorders or phonemic-based phonological disorders (Bauman-Wängler, 2004; Bernthal & Bankson, 2003), which distinguished motor- and linguistic-based speech deficits, respectively. This framework, however, neglects both the etiologic foundations of some cases of SSD and the interaction of the motor and linguistic elements of speech production (Goffman, 2005). More complex taxonomies have been proposed that categorize subgroups of SSD on the basis of etiology (Davis, 2005; Shriberg, Austin, Lewis, McSweeny, & Wilson, 1997a; Shriberg et al., 2010a) or speech sound error types (Dodd, 1995b; Dodd & McCormack, 1995). Each of these taxonomies posits a subpopulation of SSD whose disorder results from deficiencies or differences in speech motor control and coordination. Direct physiological markers for a subpopulation of this type have not been identified (Strand, McCauley, Weigand, Stoeckel, & Baas, 2013).

Figure 1 depicts the Speech Disorders Classification System (SDCS; Shriberg, Austin, Lewis, McSweeny, & Wilson, 1997b; Shriberg et al., 2010a, 2010b), an etiological classification system for SSD. In clinical typologies, the SDCS includes two classes of children with SSD: SD and motor speech disorders (MSD). Whereas the speech of children in the SD class typically normalizes by school age, with some errors persisting until approximately age 9 years, the segmental errors and prosodic and vocal features of children in the MSD class typically persist into adolescence and, for some speakers, for a lifetime (e.g., Shriberg et al., 2006). The primary speech processing deficits in two subgroups of MSD—motor speech disorders–apraxia of speech and motor speech disorders–dysarthria—are presumed to be in transcoding (planning–programming) and feedforward processing and in neuromotor execution, respectively (Shriberg & Strand, 2014). The underlying speech processing deficits and diagnostic signs of a third putative subgroup of MSD, termed *motor speech disorder–not otherwise specified* (MSD-NOS), are presently unspecified, pending empirical study (Shriberg et al., 2010a). Crucially, deficits in speech motor control in children provisionally classified as MSD-NOS are a primary risk factor for the 25% of children with SD (i.e., 4% of children overall) whose SSD persists past age 6 years (Flipsen, 2003; Goozée et al., 2007; Shriberg et al., 1999). Early identification of children with MSD-NOS may improve prognostic estimates by clinicians who treat these disorders in preschool-age children, with implications for the inclusion of a motor focus in treatment.

In this study of SSD, we included many measures of speech performance, speech acoustics, and articulatory

Figure 1. The Speech Disorders Classification System (SDCS). L. D. Shriberg, H. L. Lohmeier, E. A. Strand, & K. J. Jakielski, *Clinical Linguistics & Phonetics,* 2012; 26 (5): 445–482, copyright © 2012, Informa Healthcare. Adapted with permission of Informa Healthcare.



Downloaded From: http://jslhr.pubs.asha.org/ by a Health Sci Learning Ctr User on 04/14/2015
Terms of Use: http://pubs.asha.org/ss/Rights_and_Permissions.aspx

movement from a group of 3- to 5-year-old children with SSD ($N = 97$) to seek empirical support for subgroups within SSD. The SDCS provided an organizing framework within which discovered subgroups might be explained (see technical discussion of the term *discovered* in the next section). Clinical experience also supported the important a priori assumption that a small subgroup of children would be distinguishable from other children with SSD primarily on the basis of reduced motor performance, and the target measures selected for analysis were specifically based on this assumption. Because the differences in speech production for children with MSD-NOS might be observable in underlying speech movements, we included measures of articulatory speech kinematics for their potential to distinguish children in this putative subgroup.

Extensive study of developing speech behaviors has described the physiological framework of early and later speech acquisition (Connaghan, Moore, & Higashakawa, 2004; Goffman, 1999; Green, Moore, Higashikawa, & Steeve, 2000; Green, Moore, & Reilly, 2002; Green et al., 1997; Green & Wilson, 2006; Moore, Caulfield, & Green, 2001; Moore & Ruark, 1996; Ruark & Moore, 1997; Smith & Zelaznik, 2004; Steeve & Moore, 2009; Walsh, Smith, & Weber-Fox, 2006; Wohlert & Smith, 2002). Using palatography and electromagnetic articulography, lingual gestures produced by older children (i.e., older than 9 years) with persistent speech sound errors have been shown to be distinct from those of their typically developing peers (Gibbon, 1999; Gibbon & Wood, 2002; Goozée et al., 2007), providing support for the notion that speech motor control differences are associated with some SSD. In preschool-age children, measurements of speech movement have included point metrics (e.g., maximum displacement) as well as more dynamic measures of whole-word and phrase movements (e.g., the spatiotemporal index; Smith, Goffman, Zelaznik, Ying, & McGillem, 1995). It is not clear which, if any, of these measures would be expected to be sensitive and specific to any differences in speech movement that might exist among preschool-age children with SSD. Thus, a reasonable initial approach to identifying subgroups within a larger sample of children with SSD was to include a large number of measures of different types and from different speech and nonspeech oral behaviors, seeking convergent evidence of categorical differences. Algorithmic subgroup discovery could then be used to identify possible subgroups and the measures that best distinguish these subgroups.

Subgroup Discovery

SUBARP (Truemper, 2009), a machine learning algorithm, provided a number of distinct advantages for analyzing these data, including the output of rules for subgroups that were interpretable in the context of expert real-world knowledge. Another advantage was that, as with other subgroup discovery algorithms, a predetermined significance level could be set so that only identified subgroups exceeding this threshold were reported. Imposing a threshold introduced the possibility that we would identify no subgroups, which enhanced confidence in the reliability of the results. Alternatively, multiple significant subgroups could be discovered in a data-driven fashion, without requiring an estimate (i.e., constraint) of the number of subgroups. A particularly important feature of SUBARP in the current application was its capacity to distinguish subgroups that accounted for less than 5% of the sample (i.e., rare subgroups). Finally, given a large number of measures, SUBARP did not require a comparably large number of cases (participants) to identify relevant subgroups. This capability permitted a sample of modest size to be divided between training and testing sets for cross-validation without any associated loss in the ability of the algorithm to discover subgroups. For any subgroup discovered in the training set, confidence about the existence of that subgroup in future samples and the larger population could be estimated from the testing set. SUBARP is derived from the classification method Lsquare and related results (Bartnikowski, Granberry, Mugan, & Truemper, 2006; Felici, Sun, & Truemper, 2006; Felici & Truemper, 2002, 2005; Mugan & Truemper, 2008; Truemper, 2009) using Lavrač et al.'s (2004) approach. The Appendix includes a detailed description of the SUBARP procedures.

Research Questions

The investigation was designed to address two experimental questions in a sample of 97 preschool-age children with SSD.

1. Are there distinct subgroups of SSD that can be identified by applying a subgroup discovery algorithm to a large set of measures of auditory–perceptual, speech acoustic, and articulatory kinematic features of speech performance? Specifically, is there evidence for a subgroup of SSD that is distinguished by measures consistent with atypical speech motor control?
2. Which auditory–perceptual, acoustic, kinematic, and demographic measures best differentiate subgroups identified in the present sample of children with SSD?

Method

Participants

Ninety-seven children were enrolled in this study. Participants had receptive language skills within normal limits, as measured by a scaled score of 7 or greater on the Linguistic Concepts subtest of the Clinical Evaluation of Language Fundamentals—Preschool (Wiig, Secord, & Semel, 1992). In addition, participants had oral structures within normal limits as evaluated by the Oral/Speech Motor Control Protocol (Robbins & Klee, 1987). Participants' hearing thresholds were within normal limits on the day of testing as screened with pure-tone audiometry (25 dB HL at 1, 2, and 4 kHz). Finally, inclusionary criteria required at least one of the following for each participant:

1. Referral to the study by a certified speech-language pathologist (SLP) noting a diagnosis, based on a formal diagnostic evaluation, of a moderate to severe SSD.

Downloaded From: http://jslhr.pubs.asha.org/ by a Health Sci Learning Ctr User on 04/14/2015
Terms of Use: http://pubs.asha.org/ss/Rights_and_Permissions.aspx

2. Classification as SD using SDCS procedures, including an error profile from the 100 first-occurrence words in the child's conversational speech sample. At the time these classifications were made, the SDCS software did not include classification algorithms for MSD or any of the three subgroups shown in Figure 1.

On the basis of SDCS criteria, 61 (63%) of the 97 participants were classified as SD, 12 (12%) were classified as normal or normalized speech acquisition (NSA), and 24 (25%) were classified as between NSA and SD (NSA–SD). Children who were included solely on the basis of the SLP's judgment of moderate to severe SSD (i.e., without a congruent diagnosis using the SDCS) provided variance to the sample that was meant to reflect the true population of children being treated for SSD (i.e., most children seen clinically are not classified using the SDCS but are diagnosed using the expert opinion of a SLP).

Participants were 36 to 59 months old (M = 46 months, SD = 4 months). Consistent with prior findings of a 2:1 ratio of boys to girls with SD (Campbell et al., 2003), 66 (68%) of the children in the present sample were male. All participants were monolingual speakers of English from the Pittsburgh, Pennsylvania, metropolitan area. A questionnaire adapted from Tomblin (1989) was administered to determine the presence of a developmental communication disorder in participants' first-degree relatives. Of the sample, 34% (n = 33) had a positive family history of communication disorders, which was slightly higher than that reported in prior samples (28.1%; Campbell et al., 2003).

Tasks

Measures of performance on five speech and nonspeech tasks from a larger protocol were included in the analyses. The tasks and the rationale for inclusion are described in Table 1. Measures of nonspeech tasks (i.e., chewing and vertical jaw oscillation) were included because these measures were found to be sensitive to differences in developmental subgroupings of preschool children with typical speech (Vick et al., 2012).

Data Acquisition

During acquisition of the kinematic data, children were seated in a Rifton positioning chair fitted with a table. They were instructed to sit upright and to keep their hands on the table, holding a plush toy to avoid hand and arm movements, which might have introduced artifacts into the speech movement data. As described in Table 1, tasks used for acquisition of speech data were elicited via imitation of recorded adult female model productions. Nonverbal tasks were elicited via instructions to the child (chewing or imitation [silent vertical jaw oscillation]). To acquire the conversational speech sample, participants engaged in a play session with an experimenter without restrictions on position or movement. The play session took place before the placement of the markers for kinematic tracking. The examiners followed a standard SDCS protocol to obtain conversational speech samples about events in the participant's lives (e.g., Shriberg, Potter, & Strand, 2011).

Audio data. Audio recordings of the session were obtained using a lapel-style wireless microphone (Shure model UI-UA) affixed to the child's forehead with surgical tape. This placement provided a fixed microphone-to-mouth distance. When the child would not tolerate this placement, the microphone was taped to the headrest of the chair. The signal from the microphone was amplified using a Mackie 12-channel mixer (Model 1202-VLZ Pro). The amplified signal was recorded with a video recorder (Panasonic, AG-1980) and then filtered for antialiasing and digitized with the video signal at a sampling rate of 44.1 KHz.

Video (articulatory) data. Vertical movement records of the upper lip, lower lip, and jaw were extracted from the video recordings. An infrared camera and light source (Burle, TC351A) were used to record the movement of small (3-mm), flat, circular reflective markers attached in the midline of the child's upper lip, lower lip, and jaw (above the mental symphysis). Additional markers were placed on the tip of the nose and the bridge of the nose to provide landmarks for correction of head movement, which was accomplished algorithmically by the motion tracking software. A reference frame with two markers, 2 cm apart, was affixed to each child's forehead to calibrate distance. Each video was reviewed and logged, and task events were digitized for subsequent parsing and analysis. Two independent computer-based movement tracking systems were used to extract position in the frontal plane (i.e., vertical and lateral positions) of the markers in Cartesian coordinates from the digitized video recordings. The first was Version 6.05 of Motus (Peak Performance). The second was DS-MTT Version 2 by Henesis, a custom MATLAB routine created for movement tracking for this project, which was developed later in the data acquisition phase to improve the rate of data processing. Intersystem reliability was confirmed using 15% of the data with both systems, which yielded high concordance (>90%). The sampling rate for the kinematic data was 60 Hz. The movement records for the upper lip, lower lip, and jaw were low-pass filtered (f_{lp} = 15 Hz) forward and reverse with a digital, zero-phase shift, third-order Butterworth filter. In addition, the best straight-line linear trend was removed from each displacement record to correct for very low frequency artifact.

Data Processing and Standardizing

All parsing and transcription of the data were completed with blinding as to the diagnostic status of the participants. Data were parsed in accordance with a parallel data set from preschool-age children with typical speech acquisition described in a prior report (Vick et al., 2012); the temporal overlap of these analyses reduced potential bias arising from the researcher's knowledge of a participant's diagnosis.

Acoustic parsing. Each imitation of a target stimulus was parsed with reference to the audio signal from the video recorder. The experimenter listened to and inspected the

Downloaded From: http://jslhr.pubs.asha.org/ by a Health Sci Learning Ctr User on 04/14/2015
Terms of Use: http://pubs.asha.org/ss/Rights_and_Permissions.aspx

Table 1. Tasks completed by all participants.

Task	Description and goal	Stimuli
Conversational speech sample	Participants engaged in play session to evoke a 15-min sample. Narrow transcription of audio recording as well as error analysis of 100 first-occurrence words were completed in PEPPER (Shriberg, Allen, McSweeny, & Wilson, 2001). Provided all data for production phonology and the single measure of language production (average words per utterance).	Age-appropriate toys were used to elicit the sample
Lexical stress task	Five imitations of each of two lexical stress (trochaic and iambic) bisyllables in CVCV context. Six words (*baba, mama,* and *papa* in trochaic and iambic stress) were produced, with five repetitions of each bisyllable imitated in a row. Provided perceptual, acoustic, and kinematic data for each production as well as measures of acoustic and kinematic variability on multiple repetitions.	Recorded adult female model
Nonword repetition task	Four words from the Syllable Repetition Task (Shriberg et al., 2009). The nonwords were *bada, bama, bamana,* and *manaba*, produced with equal stress. Five repetitions of each target production were presented to each participant, alternating with the participant's imitation of the model. The same token was imitated five times before progressing to the next token. Provided information about speech processing in increasingly complex contexts (two- and three-syllable nonwords). Contained only four of the Early 8 consonants and a single low back vowel. Multiple repetitions of each nonword allowed for measurement of acoustic and kinematic variability.	Recorded adult female model
NS tasks	The NS tasks included two trials of chewing a cracker and five trials of silent vertical jaw oscillations. The tasks provided measures of maximum mandibular displacement in NS context as well as measures of NS cyclic kinematic variability.	Live adult, female model

Note. PEPPER = Programs to Examine Phonetic and Phonological Evaluation Records; NS = nonspeech.

waveform of each segment and roughly parsed the onset and offset of the entire production using a graphical user interface of a custom-scripted MATLAB algorithm. The algorithm subsequently added 50 ms to the beginning and end of the parsed signal to ensure inclusion of complete acoustic information for later perceptual judgments.

The vowels of individual syllables were then closely parsed for acoustic analyses. The vowel onset was defined as the first positive-going zero crossing in the signal when the waveform became periodic (i.e., vocalic); the offset was defined as the final negative-going zero crossing in the periodic signal associated with the vowel. The audio record was accompanied by a trace of the signal amplitude. To automate parsed landmarks, an amplitude threshold at 15% of the maximum amplitude produced was overlaid on the amplitude envelope. The intersection of the amplitude envelope and the 15% threshold was used to identify the beginning and end of each vowel. User-selected landmarks were "snapped" to this intersection by the algorithm. Audio playback assisted users in completing fine acoustic parsing. Panel A of Figure 2 depicts the acoustic parsing user interface.

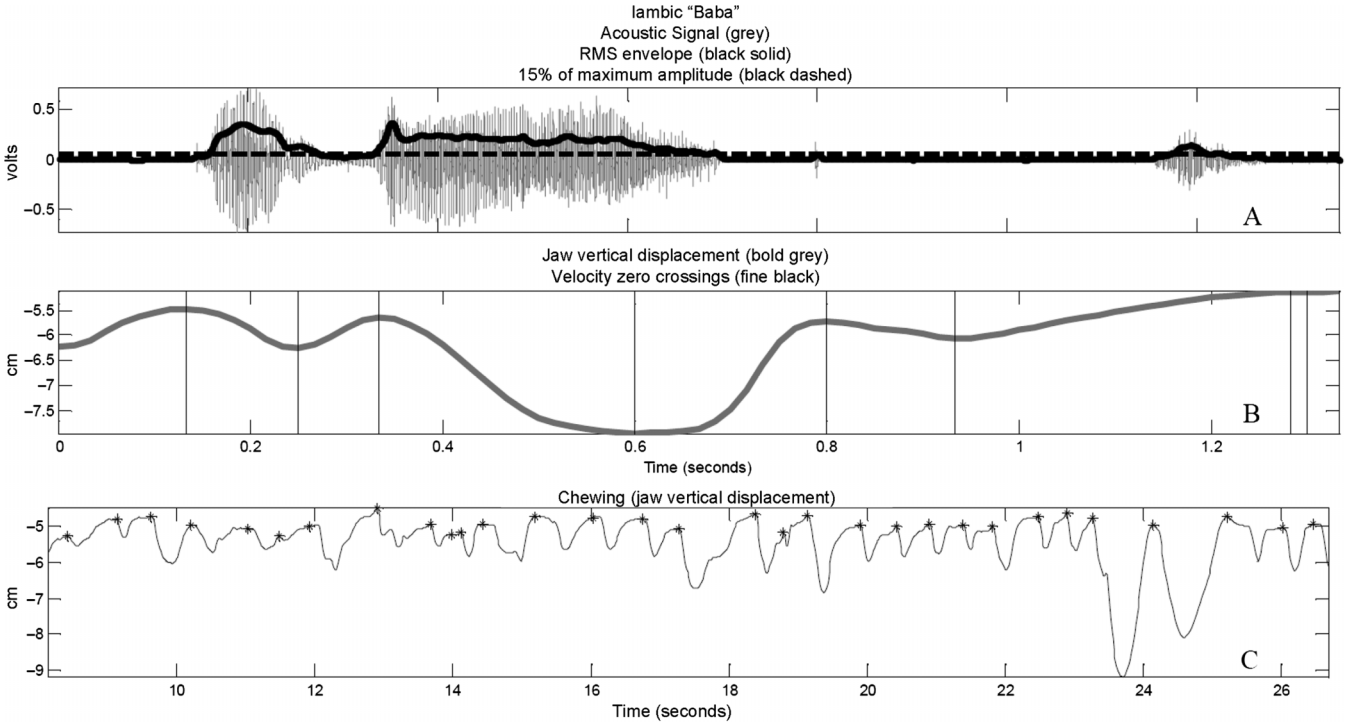
Jaw, upper lip, and lower lip kinematic parsing. For the speech tasks, movement trajectories were obtained for the markers on the upper lip, lower lip, and jaw. Velocity zero crossings in the vertical jaw displacement record were used to parse the onset and offset boundaries of all three (upper lip, lower lip, and jaw) position traces (Green et al., 2000). The velocity of vertical jaw position was derived from the position record, and zero crossings were displayed as vertical lines over the displacement record. The onset boundary was operationally defined as the last negative-traveling zero crossing in the velocity waveform before jaw depression for the vowel; the first velocity zero crossing during jaw depression for the final syllable was used to mark the offset boundary. The time indices for these onset and offset boundaries were used to parse the displacement trajectories for the upper and lower lip. Panel B of Figure 2 depicts the speech task kinematic parsing interface.

Because the movement of the jaw contributes substantially to the displacement of the lower lip, the displacement of the marker on the lower lip represented the combined movement of the lower lip and the jaw. Accordingly, the jaw displacement signal was subtracted, sample by sample, from the lower lip displacement signal. The resulting trajectory was the record used to represent lower lip movement (Green et al., 2000).

Nonspeech task parsing. For the nonspeech tasks, measures were based only on the position trace from the jaw. Chewing or vertical jaw oscillation samples that had fewer than three cycles or exhibited movement artifact were excluded from the analyses. First and last chewing cycles were removed from each chewing trial as well. Because the measures for these tasks included cycle-to-cycle measures, each cycle was demarcated. Jaw elevation–depression–elevation (open–close) cycles were parsed algorithmically, marking each cycle boundary at its peak elevation (identified by the associated zero velocity point). Because of the irregular displacement signal associated with molar contact during chewing, numerous zero crossings in the velocity record occurred during some instances of jaw elevation. In these cases, the algorithm specified the rightmost zero crossing as the cycle onset–offset boundary. Occasionally, the algorithm would yield a false positive or false negative boundary selection, in which case the user was able to modify the selections in the interface by either adding or subtracting

Downloaded From: http://jslhr.pubs.asha.org/ by a Health Sci Learning Ctr User on 04/14/2015
Terms of Use: http://pubs.asha.org/ss/Rights_and_Permissions.aspx

Figure 2. Panels A and B plot acoustic and kinematic data, respectively, for an iambic production of *baba*. In Panel A, the amplitude envelope is plotted in solid black, and 15% of the maximum amplitude is plotted as a dashed line. The beginning and end of vowels were parsed algorithmically at the intersection of these lines. Panel B is a plot of the vertical displacement of the jaw marker for the same production. The vertical lines are the velocity zero crossings used to parse the kinematic signals. Panel C shows the vertical displacement of the jaw of the same participant while chewing a cracker. Asterisks mark the closing landmarks that were selected algorithmically for the trial. RMS = root-mean-square.



points. Panel C of Figure 2 depicts the non–speech task kinematic parsing interface.

Spontaneous speech sample transcription and analysis. All data acquisition, data reduction, and data analyses of the conversational speech sample used well-developed procedures for research in pediatric SSD (Phonology Project Laboratory Manual; unpublished). Narrow phonetic transcription of the continuous speech samples was completed by two experienced transcribers using procedures described by Shriberg et al. (2010b). The procedure included perceptual use of diacritics sensitive to articulatory place, manner, voicing, duration, and force. For instance, the check symbol in the Clinical Phonetics system of diacritics was used to indicate weakened articulatory force, most frequently for weakly ploded voiceless stop consonants, which are common signs in structural (e.g., velopharyngeal incompetence) and motor (e.g., subtypes of dysarthria) SSD. In a study of transcription reliability for children with SD that included 10 children from the present study, point-to-point interjudge and intrajudge agreement were 86.7% and 91.8%, respectively (Shriberg et al., 2010b). In that study, some estimates of consonant and vowel transcription agreement were in the mid-60% range, values that have also been reported in prior estimates of broad and narrow phonetic transcription agreement (McSweeny & Shriberg, 1995; Shriberg et al., 1997a; Shriberg & Lof, 1991; Shriberg et al., 2005). The Shriberg et al. (2010b) estimate referenced previously reported that agreement for narrow transcription averaged 6.1% lower than agreement for broad phonetic transcription. One agreement estimate for the transcription system used in the present study indicated that the standard error of measurement was approximately 4% (Shriberg et al., 1997b). Thus, there are commonly acknowledged constraints on speech data from phonetic transcription; see Shriberg et al. (2010b) for comparable constraints on the reliability of some types of acoustic data.

Perceptual analyses. Perceptual analyses of the lexical stress task and the nonword repetition tasks were completed using the parsed audio signals. In the lexical stress perceptual task, audio files for each child's productions of each of the bisyllables were presented in randomized order in blocks of 10 participants (i.e., about 300 audio files in each test block). Two listeners (graduate students in speech-language pathology) assessed whether each item was produced with the intended phonemic target and identified which syllable was stressed (first, second, or both, when even stress was perceived). All 2,617 contrastive stress productions were judged by two listeners. Joint probability concordance between the two listeners was 89.1%. As can be seen in Table 2, average phonemic accuracy was comparable for trochees (78.8%; 1,066/1,352) and iambs (73.8%; 934/1,265); however, average accuracy of imitative stress

Downloaded From: http://jslhr.pubs.asha.org/ by a Health Sci Learning Ctr User on 04/14/2015
Terms of Use: http://pubs.asha.org/ss/Rights_and_Permissions.aspx

Table 2. Perceptual analysis of lexical stress and nonword repetition tasks.

Type and word	Attempted, *n*	Phonemes correct		Stress correct		Overall	
		n	% of attempted	*n*	% of attempted	*n*	% of attempted
Trochee							
baba	447	350	78.3	358	80.1	304	68.0
mama	446	366	82.1	373	83.7	335	75.1
papa	459	350	76.3	356	77.6	289	63.0
Total	1,352	1,066	78.9	1,087	80.4	928	68.6
Iamb							
baba	427	315	73.8	269	63.0	233	54.6
mama	423	333	78.7	245	57.9	223	52.8
papa	415	286	68.9	245	59.0	193	46.5
Total	1,265	934	73.8	759	60.0	649	51.3
Nonword							
bada	476	366	76.9	NA	NA	345	72.5
bama	466	266	57.1	NA	NA	246	52.8
bamana	468	169	36.1	NA	NA	158	33.8
manaba	444	113	25.5	NA	NA	99	22.3
Total	1,854	914	49.3	NA	NA	848	45.7
Grand total	4,471	2,914	66.2	1,846	70.5	2,425	54.2

Note. Includes number attempted and judged accuracy of phonemics and stress. *Overall* refers to the productions that were suitable for all measures in the analysis, which required correct phonemics and stress, as well as viable kinematics (i.e., without movement artifact). NA = not applicable.

was better for trochees (80.4%; 1,087/1,352) than for iambs (60.0%; 759/1,265). The phonemic accuracy of the nonword repetition task productions was judged blindly (i.e., to production, participant, and diagnostic status) using an open-set, broad phonemic transcription task. No judgment of the lexical stress of these productions was completed. The items were presented to listeners randomized across participants and production types. Three listeners judged each production. Transcription decisions used a best-two-of-three criterion (i.e., agreement by at least two of the three judges). This criterion was reached for 95.2% (1,765/1,854) of the productions. For the remaining productions, the raw video file was reviewed by a single judge to finalize transcription. Forty-nine percent (914/1,854) of the nonwords were judged to be phonemically correct. Phonemic accuracy decreased through the experimental paradigm from the first-attempted, two-syllable productions to the final three-syllable productions, with a drop in accuracy from the two-syllable (88.1%; 616/699) to the three-syllable (30.9%; 282/912) productions (*bada,* 76.8% [366/476]; *bama,* 57.1% [266/466]; *bamana,* 36.1% [169/468]; and *manaba,* 25.4% [113/444], in the order produced in the protocol). Details for phonemic accuracy are reported in Table 2.

Measures

Fifty-three continuous and three categorical variables were processed for inclusion in the analysis. The measures sampled four levels of observation, including auditory–perceptual, acoustic, kinematic, and demographic domains. This broad range of observations was predicated by the need to avoid a priori assumptions of the number or types of groups that might emerge from the analysis. Each domain included a number of redundant measures for each type of speech production (i.e., two- or three-syllable productions), which yielded overlapping samples of similar measures from verbal productions across levels of difficulty. For some participants, for instance, behavioral performance was similar for both two- and three-syllable productions, whereas for others, performance decreased substantially with the added complexity of a three-syllable imitation. In addition, although numerous measures have been reported, we did not know which of the contributing measures would be most effective in identifying emergent groups of preschool children with SSD. Descriptions of each measure can be found in the supplemental materials. Means and standard deviations for each measure are reported in Supplemental Table 1.

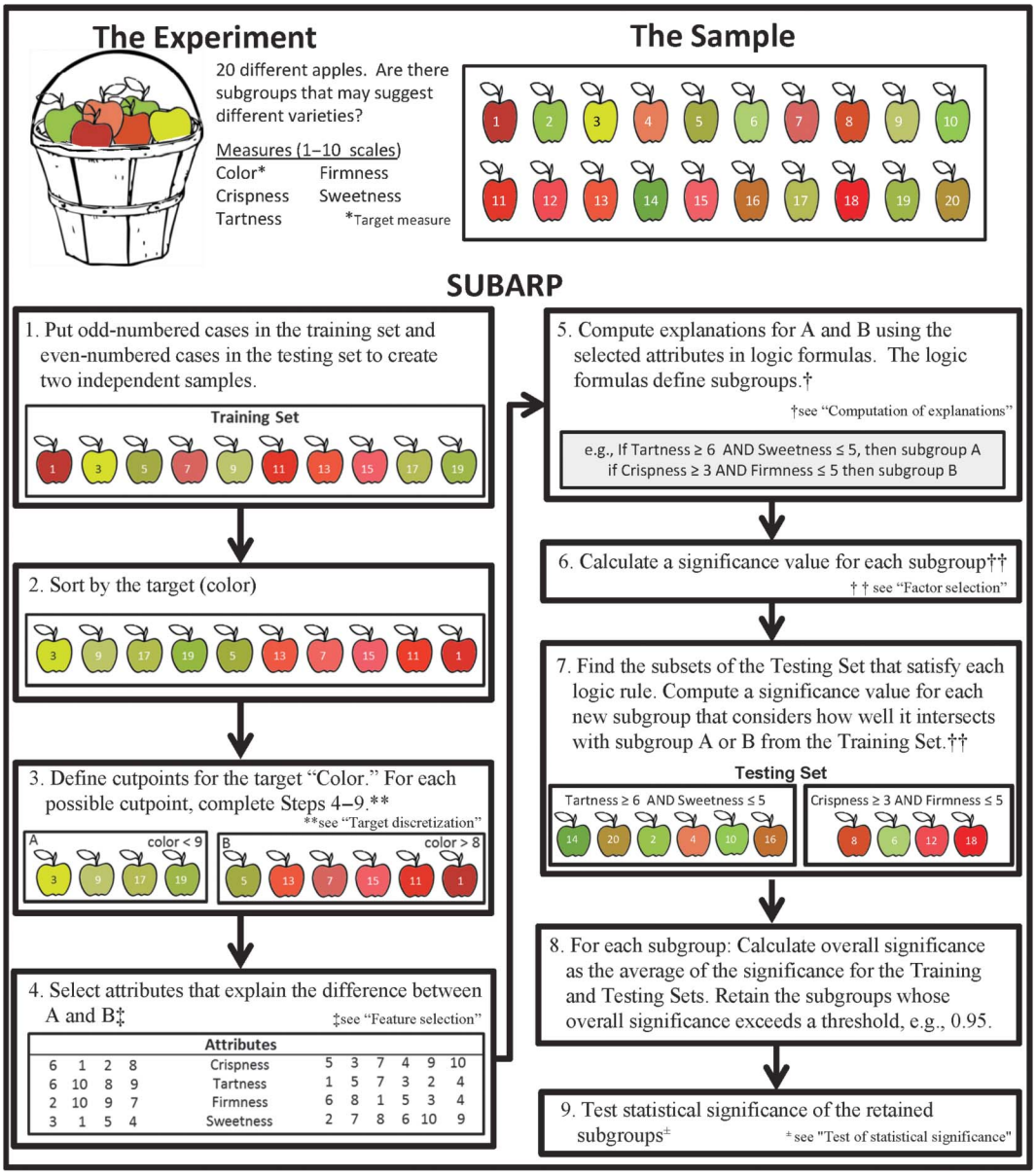
Subgroup Discovery

To identify previously unknown subgroups, the entire data set was input to the SUBARP algorithm. Details regarding the algorithm can be found in the Appendix, and a schematic of the algorithm using a simple data set can be found in Figure 3. Each participant in the present data set was assigned a case number, consecutively from the date of admission to the study. To initiate SUBARP, odd-numbered cases were put in the training set and even-numbered cases were assigned to the testing set (see Step 1 in Figure 3), with a net of 49 cases in the training set and 48 cases in the testing set.

Targets and attributes. Whereas only one of the five measures in the example in Figure 3 was assigned as a target (i.e., color), 34 of the 53 measures from the data were assigned as targets and the remaining 19 measures were designated nontarget attributes. In SUBARP, target measures become the primary defining attributes of any discovered subgroup. For this reason, only measures that would be of

Downloaded From: http://jslhr.pubs.asha.org/ by a Health Sci Learning Ctr User on 04/14/2015
Terms of Use: http://pubs.asha.org/ss/Rights_and_Permissions.aspx

Figure 3. Schematic of the SUBARP algorithm that documents the steps taken to discover subgroups. In the schematic, apples are used as a simple example. A full explanation of SUBARP can be found in the Appendix. Specific sections of this explanation are referenced in each section of the schematic.



interest as defining attributes are selected as targets. For instance, in this experiment in which a subgroup was sought whose members had deficits in speech motor control, measures of articulatory variability were of interest as targets, whereas demographic measures, such as age, were not. To further refine the model, we only included as targets measures for which we had data from the largest number of cases. For example, because most children in the sample produced two-syllable tokens that could be analyzed, measures of performance on two-syllable productions were included as targets. Three-syllable tokens, which were not produced by as many children, were submitted to SUBARP as nontarget attributes. For highly correlated measures that resulted from the Programs to Examine Phonetic and Phonological Evaluation Records (PEPPER) analysis (e.g., percentage of consonants correct and revised percentage of consonants correct) the revised version was selected as the target. As defined in the Appendix, the revised versions do not count distortions or allophonic variations as errors and are more sensitive to differences in development and diagnostic groups (Campbell, Dollaghan, Janosky, & Adelson, 2007; Shriberg et al., 1997b).

Downloaded From: http://jslhr.pubs.asha.org/ by a Health Sci Learning Ctr User on 04/14/2015
Terms of Use: http://pubs.asha.org/ss/Rights_and_Permissions.aspx

Each target measure was run through the SUBARP algorithm iteratively as detailed in Figure 3. During each iteration, other target measures served as nontarget attributes from which logic rules could be generated for each subgroup.

Running SUBARP. Discretization of each target occurred after SUBARP sorted the records in ascending order for the value of the target (see Step 2 in Figure 3). For the discretization of targets, SUBARP was set to use as many as 50 cut-points for each target. Given 49 training records, this meant that SUBARP evaluated all possible cut-points for each target. SUBARP created and evaluated 1,209 subgroups, each being defined by a rule involving a target measure and two modifying rules involving attribute measures (see Steps 4 and 5 in Figure 3). Preliminary testing suggested that rules with one target and two attributes were most interpretable. SUBARP used a compound measure of significance that calculated (a) the proportion of cases that fulfilled the target rule that also fulfilled the additional attribute rules and (b) the likelihood that a similar group could be generated using a random process (see Step 6). Each significance value was calculated twice, once for the training data and once for the testing data (see Step 8). The average of the resulting two significance values was the overall significance of the subgroup. Subsequently, a binomial probability test was completed, resulting in a value ranging from 0 to 1 (see Step 9). This test determined the extent to which the target and attribute variables were related. Smaller values suggested closely related target and attribute values. For details of these computations, see the Appendix.

Selection of SUBARP solutions. Only solutions with an overall significance of .95 or greater were retained, which resulted in 13 potential subgroups for the present data. Next, 11 subgroup solutions were eliminated from consideration because they intersected with other solutions (i.e., same targets and attributes but different cut-points) and were less significant; although these solutions were statistically significant, they were redundant permutations of the resulting two subgroups. Thus, the target and attribute rules in Table 3 defined the remaining two subgroups, which made up the results of the SUBARP analysis for the present data. The significance levels obtained for each of the two subgroups (.95 for Subgroup A and .99 for Subgroup B) suggested that each rule set identified a distinct subgroup (i.e., <5% probability of finding these groups by chance). The significance values (i.e., $p < .05$) from the binomial probability test suggested a high degree of interrelatedness between the target and attribute values specified in the rules, meaning that the combined attributes had high predictive value for the targets.

Table 3. Target and attribute (measures) rules defining each subgroup and overall significance after subgroup identification using both the training and testing sets.

Group and target–attribute rules	Value
Subgroup A	
Target: Proportion of tasks attempted	>.72
Attribute 1: Proportion of tasks with correct phonemes	>.36
Attribute 2: Proportion of iambic targets imitated with correct stress	>.17
Overall significance	.95
Binomial probability	.0006
Subgroup B	
Target: Proportion of iambic targets imitated with correct stress	<.17
Attribute 1: Proportion of tasks attempted	<.72
Attribute 2: Mean upper lip maximum displacement (cm)	>.20
Overall significance	.99
Binomial probability	.036

The proportions of participants in the training and testing sets of each subgroup were compared with a two-proportion z test to confirm that these split samples were not statistically different and that their sizes provided a reliable estimate of population prevalence. To estimate the range of population prevalence for the two subgroups, 95% confidence intervals (CI) were calculated on the overall proportion of participants in each group.

Statistical analyses of subgroups. The SUBARP rules for membership specified the level of performance on six measures that was necessary and sufficient to be included in one of the two emergent subgroups. Performance on the original 53 continuous measures by members of the two subgroups was compared using independent t tests. To adjust the alpha level for multiple comparisons, the threshold for statistical significance was set at $p < .0007$. The proportions of each subgroup falling within each of the categories of the dichotomous qualitative measures (i.e., sex and family history) were compared using chi-square tests. Forward-stepping discriminant analysis (DA) was used to characterize and graphically represent the two groups using the 47 continuous behavioral, acoustic, and kinematic measures that were not the target and attribute measures used to identify the subgroups with SUBARP. Thirteen participants were not classified in either of the two subgroups. These participants' performance was also graphically represented using the DA.

Results

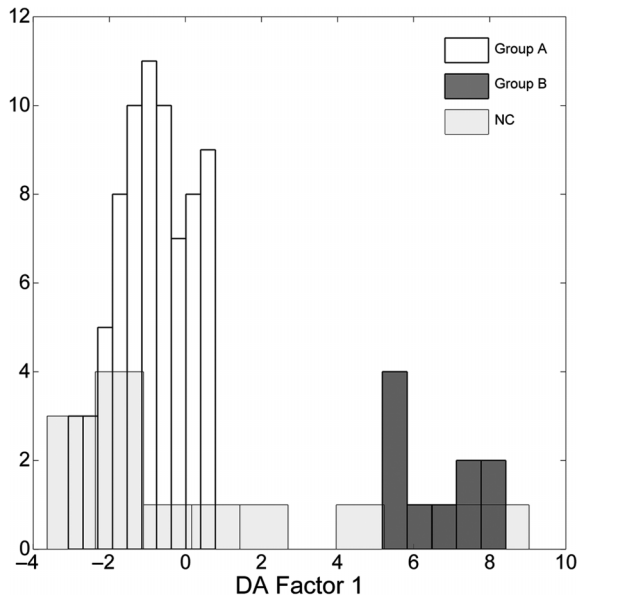
Two subgroups emerged from the SUBARP analysis as statistically significant. The first subgroup, termed *Group A* ($n = 74$), consisted of the majority (76.2%) of participants in the sample, and the second subgroup, termed *Group B* ($n = 10$), made up 10.3% of the sample. A total of 13 participants (13.4%), termed *not classified* (NC), were not classified using the SUBARP routine. Supplemental Table 1 includes group-based performance data for each of the original 53 continuous measures and Table 3 provides the target and attribute rules that defined both subgroups.

DA was used to evaluate pairwise differences between Groups A and B on each of the 53 continuous measures; this analysis yielded a single linear DA factor. The DA factor in this approach is the linear combination of measures that provided the best separation between members of Group A and members of Group B. Each contributing measure from the model has an associated correlation with the DA factor so that it is possible to identify which measures best discriminated between the two groups.

Downloaded From: http://jslhr.pubs.asha.org/ by a Health Sci Learning Ctr User on 04/14/2015
Terms of Use: http://pubs.asha.org/ss/Rights_and_Permissions.aspx

Significant correlations between the DA factor and the continuous measures are presented in Table 4. Figure 4 depicts the performance of each participant on the continuum of the DA factor. Participants with a high DA factor score (Group B participants) had high scores on the measures that had a positive correlation with the DA factor; participants with a low DA factor score (Group A participants) had high scores on the measures that had a negative correlation with the DA factor. For example, participants in Group A had a high proportion of tasks with accurate phonemics and no kinematic artifact, and the associated productions tended to be produced with correct lexical stress and phonemics; these measures were negatively correlated with the DA factor. These associations permitted more detailed descriptions of the individual group characteristics.

Figure 4. Discriminant analysis (DA) results: Selected measures that were positively correlated with Factor 1 appear on the right side of the graph, and selected measures that were negatively correlated with Factor 1 appear on the left side of the graph. NC = not classified.



Group A Characteristics

Of the 74 participants in Group A, 48 (64.9%) were male, which was not significantly different from the proportion of boys in the entire sample, $\chi^2(1) = 0.48$, $p = .49$. The mean age of Group A members was 46.7 months (Index 52 in Supplemental Table 1), which was not statistically different from that of Group B (42.5 months), $t(82) = -1.996$, $p = .049$. Of the total sample of 97 children, 33 (34%) had a positive family history of communication disorders; 21 of the members (28.4%) of Group A had a positive family history of communication disorders, which did not differ significantly from the percentage in the whole sample, $\chi^2(1) = 1.04$, $p = .31$. Of the total sample of 97 children, 13 (13.4%) were classified as NSA by the SDCS. Ten children in Group A were classified as NSA (13.5%); this relative proportion was independent of group membership (i.e., Group A or B), $\chi^2(4) = 3.77$, $p = .44$. As predicted by the rules in SUBARP that defined Group A membership, and compared with Group B, individuals in Group A attempted more of the target tasks (Index 1), $t(82) = -22.57$, $p < .0001$; produced more phonemes accurately (Index 2), $t(82) = -8.65$, $p < .0001$; and imitated iambic stress with greater accuracy (Index 4), $t(82) = -9.66$, $p < .0001$. Performance for this group also exceeded that of Group B on the other measures of task accuracy, including imitation of trochaic stress (Index 3), $t(82) = -6.1$, $p < .0001$, and the overall proportion of tasks that were produced without phonemic errors and with usable kinematics (Index 5), $t(82) = -6.88$, $p < .0001$. Most strikingly, the kinematic variables were a distinguishing characteristic of Group A compared with Group B.

The proportion of participants in the testing set who met the performance criteria for Group A was not significantly different from that of participants in the training set ($z = 1.18$, $p = .99997$), which supported the suggestion that the overall sample proportion (76.3%) was a good estimate of the population prevalence of children with SSD with characteristics like those of Group A. The estimated 95% prevalence range for children with SSD whose

Table 4. Significant positive and negative correlations of specific measures with discriminant analysis (DA) factor.

Measure	DA factor *r*	Overall, *M* (*SD*)	Group A, *M* (*SD*)	Group B, *M* (*SD*)
Proportion with no phonetic errors and good kinematics (Index 5)	–.654	0.54 (0.23)	0.63 (0.17)	0.28 (0.21)
Proportion trochees with correct stress (Index 3)	–.603	0.80 (0.26)	0.88 (0.15)	0.47 (0.44)
Acoustic variability of iambic stress marking (Index 9)	–.290	43.21 (20.04)	42.97 (17.88)	27.98 (15.49)
Age (Index 52)	–.233	46.02 (6.36)	46.70 (6.30)	42.60 (5.80)
Intelligibility Index (Index 50)	–.227	88.42 (11.28)	90.12 (10.00)	83.52 (6.38)
Trochees—jaw convergence index (Index 35)	.258	22.78 (4.90)	22.43 (4.51)	25.81 (4.42)
Variability of lower lip maximum displacement (Index 18)	.278	38.99 (10.61)	37.06 (7.86)	44.36 (15.34)
Two-syllable word duration (Index 10)	.318	11.27 (3.24)	11.15 (2.52)	12.22 (5.12)
Variability of upper lip maximum displacement (Index 17)	.429	43.49 (15.07)	40.02 (12.85)	57.30 (14.90)
Variability of jaw maximum displacement (Index 19)	.511	34.87 (10.35)	32.14 (7.64)	45.99 (13.38)

Note. $p < .05$. Index numbers refer to Supplemental Table 1, which provides descriptive statistics for each measure.

Downloaded From: http://jslhr.pubs.asha.org/ by a Health Sci Learning Ctr User on 04/14/2015
Terms of Use: http://pubs.asha.org/ss/Rights_and_Permissions.aspx

characteristics would be consistent with those of Group A is 67.8%–84.8%.

Group B Characteristics

Of the 10 participants in Group B, 8 (80%) were male; five (50%) had a positive family history of communication disorders. Neither finding was significantly different from the overall sample percentages of 68% (66/97) and 34% (33/97), respectively, nonsignificant (*ns*) $\chi^2(1) = 0.66$, and *ns,* $\chi^2(1) = 1.14$. The mean age of the children in this group was 42.5 months (*ns*). None of the children in Group B was identified as NSA by the SDCS. In addition to Group B's significant differences from Group A on task performance, a number of other significant differences in Group B's articulatory kinematics performance were observed. The SUBARP rules that distinguished members of this group identified mean maximum displacement of the upper lip (0.23 cm) as significantly greater during speech tasks for members of Group B than for members of Group A (0.19 cm; Index 14), $t(82) = 4.58$, $p < .0001$. The coefficient of variation for measures of maximum displacement of both the upper lip and jaw during speech tasks was significantly greater for Group B participants than for Group A participants, upper lip (Index 17), $t(82) = 3.91$, $p < .0001$, and jaw (Index 19), $t(82) = 4.86$, $p < .0001$. Higher values on these measures indicated greater articulatory variability.

Also of note when describing salient characteristics of children in Group B are the measures found to be significantly correlated with the DA factor that distinguished children in Group B from children in Group A, as listed in Table 4. Many of these measures were not found to be significantly different between groups in the analysis of variance but may be important diagnostic markers. Children in Group B were found to produce trochaic stress with less accuracy (proportion correct = 0.47) than children in Group A (0.88; Index 3), $t(82) = -6.10$, $p < .0001$, and marked productions with iambic stress with less acoustic variability. Children in Group B were slightly younger than children in Group A (Group B, 42.6 months; Group A, 46.7 months; Index 52), $t(82) = -2.00$, $p = .049$. In addition, the overall percentage of intelligible words in the conversational speech sample was lower for children in Group B (83.5%) than for children in Group A (90.1%; Index 50), $t(82) = -1.94$, $p = .056$. Jaw movement for bisyllables with trochaic stress was produced with greater word-level variability when productions in error were included (see Convergence Index in Supplemental Materials) for children in Group B (25.8) than for children in Group A (22.4; Index 35), $t(82) = 2.22$, $p = .029$. Finally, two-syllable word duration was greater in children in Group B (0.85 s) than in children in Group A (0.77 s; Index 10); $t(82) = 2.79$, $p = .007$.

The proportions of participants assigned to Group B in the training and testing sets were not significantly different ($z = -.298$, $p = .99998$), suggesting that the overall sample proportion (10.3%; 95% CI [4.3%, 16.46%]) was a good estimate of SSD population prevalence for children with characteristics consistent with Group B.

Not Classified

Thirteen participants (13.4%) were not classified into either Group A or Group B using the two solutions selected from SUBARP. Of the NC group, 10 participants were male (77%), *ns,* $\chi^2(1) = .48$, and six had a positive family history of communication disorders (46%), *ns,* $\chi^2(1) = 2.28$. Because those children who were not classified would not have any expected homogeneity, planned comparisons of this group with the other subgroups were not made; performance on each of the continuous measures is reported in Supplemental Table 1. Figure 4 displays the NC group relative to the other two groups on the DA factor continuum that maximally separated Groups A and B. The NC participants were uniformly distributed along this continuum.

Discussion

This investigation was designed to answer two questions. The first—whether there is statistical support for a subgroup of children with SSD who are distinguished by measures consistent with atypical speech motor control—was affirmed by the results. Two reliable subgroups were identified within the sample of 97 children with SSD. The first, Group A, consisted of a proportionally large ($n = 74$; 76%; 95% CI [67.8%, 84.8%]) subgroup of the participants, whereas the second, Group B, consisted of a substantially smaller ($n = 10$; 10.3%; 95% CI [4.3%, 16.46%]) number of participants. Children in Group B were distinguished from children in Group A by a number of measures, including some that suggested atypical speech movement, providing preliminary empirical support for a small SSD subgroup with motor speech involvement, as has been proposed by a number of investigators (e.g., Davis, 2005; Dodd, 1995a). Crucially, this subgroup could not be distinguished with measures that would typically be used in a clinical setting; moreover, the differences in speech motor control observed in this group would be subclinical. The second research question asked which measures would best differentiate any identified subgroups in the sample. The results provide a number of candidate measures that are likely to be useful in distinguishing the children in the two groups.

Groups A and B

One goal of this study was to determine whether any emergent subgroups would be consistent with characterizations of SSD subgroups proposed in the SDCS (Figure 1; Shriberg et al., 2011). The features of the two identified groups do appear to be consistent with the SD and MSD-NOS subgroups described in the SDCS (Shriberg et al., 2010a). Specifically, the majority (76%) of sample participants were identified as Group A. A post hoc comparison between the performance profiles of Group A participants and those of a group of typically developing children described previously (Vick et al., 2012) yielded no differences between the groups across the measures reported in this

Downloaded From: http://jslhr.pubs.asha.org/ by a Health Sci Learning Ctr User on 04/14/2015
Terms of Use: http://pubs.asha.org/ss/Rights_and_Permissions.aspx

study, except that participants in Group A had more speech sound errors using conversational speech measures, such as the revised percentage of consonants correct.[1] Of particular interest was that, as in children with typical speech acquisition, participants in Group A had typical motor speech skills as measured by speech and nonspeech articulatory kinematic variables. Collectively, this evidence suggests that Group A participants belong to the proportionally large SD class of the SDCS, who have no measurable speech motor involvement.

In contrast, the relatively small number of participants (10.3%) in Group B had comparably poor speech motor control, as evidenced by significantly higher articulatory variability on measures of upper lip, lower lip, and jaw movement during repeated productions of two- and three-syllable tokens. They also exhibited larger upper lip displacements. Behaviorally, participants in Group B made fewer attempts at the target tasks and produced fewer accurate phonemes and less accurate lexical stress. The low prevalence, increased speech motor variability, and behavioral characteristics of Group B are consistent with the putative MSD subgroup termed MSD-NOS in the SDCS, which includes children with suspected motor speech disorders who do not meet the criteria for childhood apraxia of speech or dysarthria (Shriberg, Lohmeier, Strand, & Jakielski, 2012). From the results of the present study, the estimated population prevalence of this subgroup is 4.3%–16.46%. These are children with the potential to be at greatest risk for persistent SSD (Shriberg et al., 2011). Future studies should study children with characteristics of Group B longitudinally to confirm the validity of this hypothesis.

The identification of this motor speech group using nonprimary features of SSD, such as speech movement variability, supports the finding that direct measures of the primary features of developmental speech disorders are not sufficient to differentiate clinically distinct populations (Connaghan & Moore, 2013). For instance, participants in both groups scored similarly on the measures resulting from the PEPPER analyses of the conversational speech samples (e.g., percentage consonants correct: Group A, 72%; Group B, 70.3%), demonstrating that children in Group B would be difficult to identify using only conventional measures of speech competence. Behaviorally, children in Group B struggled with performing the tasks in the present research protocol, which included imitating audio-recorded models of nonsense words. They were less likely to attempt the tasks (Group A, 98% attempted; Group B, 39% attempted) and, when they did, were likely to produce the nonsense words with speech sound errors (Group A, 74% correct; Group B, 31% correct), despite the composition of the models, which included only the "Early 8" consonants /b/, /p/, /m/, and /n/ and the vowel /ɑ/ (i.e., many other nonsense word tasks include diphthongs; Shriberg et al., 1997b). Relative to conversational speech, this comparatively poor performance by children in Group B during the protocol may reflect deficits in one or more of several speech processing tasks, including auditory–perceptual encoding, transcoding, and execution during productions of the imitated segments (Shriberg et al., 2012). Similarly, children in Group B were less accurate than those in Group A when imitating iambic stress (Index 4; Group A, 72% accurate; Group B, 2% accurate) and had significantly longer two-syllable durations (Index 10; Group A, 0.77 s; Group B, 0.85 s), suggesting a relative weakness in encoding suprasegmental elements (Ziegler, Staiger, & Aichert, 2010). Taken together, the findings for children in Group B are consistent with a subtype of SSD with behavioral characteristics that suggest motor speech involvement. These behavioral characteristics were corroborated by physiological markers that confirmed subclinical deficits in speech motor control that are not able to be identified with standard clinical measures.

Upper lip maximum displacement was found to be larger in Group B than in Group A. This was an interesting finding, given that no difference was found for this group in maximum displacement of either the jaw or the lower lip. One potential explanation for the difference in upper lip maximum displacement could be the unusual and often inconsistent articulatory postures observed clinically in children with suspected motor speech disorders underlying their SSD. Subjectively, three of these children were observed to have noticeable upper lip movement during speech when the videos were reviewed for marker tracking. It could well be that unusually large upper lip displacements are a salient characteristic for at least some children whose SSD results from underlying differences in speech motor control. Maximizing upper lip displacements could be used as a strategy for achieving a lip aperture goal if other degrees of freedom, such as jaw elevation, are being minimized in a pathological speech production system.

Given these findings, it was somewhat surprising to note that a number of measures of word-level variability, as measured by the spatiotemporal index, did not statistically differentiate Group B from Group A; the spatiotemporal index has been demonstrated to be a robust and sensitive indicator of differences in performance among diagnoses (e.g., specific language impairment; Goffman, 1999). Other measures of speech movement variability, specifically the coefficient of variation of maximum displacement of the upper lip, lower lip, and jaw, were found to be critical to the separation of Groups A and B with post hoc DA, demonstrating that children in Group B produced movements with significantly greater variability than those of children in Group A. This speaks to the exploratory nature of this approach and to the power of subgroup discovery for this application. To identify the critical differences between children in the two groups, it was essential to look for differences at the syllable level where maximum

[1]All of the measures from the PEPPER analysis were significantly different between the children in Group A and the children with typical speech acquisition from the Vick et al. (2012) study ($p < .0001$), with the typical children scoring higher on all of the metrics. None of the other measures were significantly different between the two groups with the exception of proportion phonetics correct (Index 1), which approached significance ($p = .03$; Group A, $M = 0.74$; typically developing children, $M = 0.80$).

Downloaded From: http://jslhr.pubs.asha.org/ by a Health Sci Learning Ctr User on 04/14/2015
Terms of Use: http://pubs.asha.org/ss/Rights_and_Permissions.aspx

displacement was measured, as opposed to the word level. It remains to be seen how these children would perform with a phrase-level measure of kinematic variability.

SDCS classification (i.e., SD vs. NSA) was not significantly different between the two identified groups. It is unlikely that including children classified as NSA using SDCS criteria (i.e., in the present context, likely consistent with normalized speech because there were no reliable age-inappropriate deletions or substitutions in continuous speech) confounded the results of the analysis, although the largest proportion of NSA children included in the analysis were classified into Group A. The implication of this finding is that SLPs typically use a number of speech criteria to classify children as having SD other than those used by the SDCS. These other factors are unlikely to be associated with differences in speech motor control.

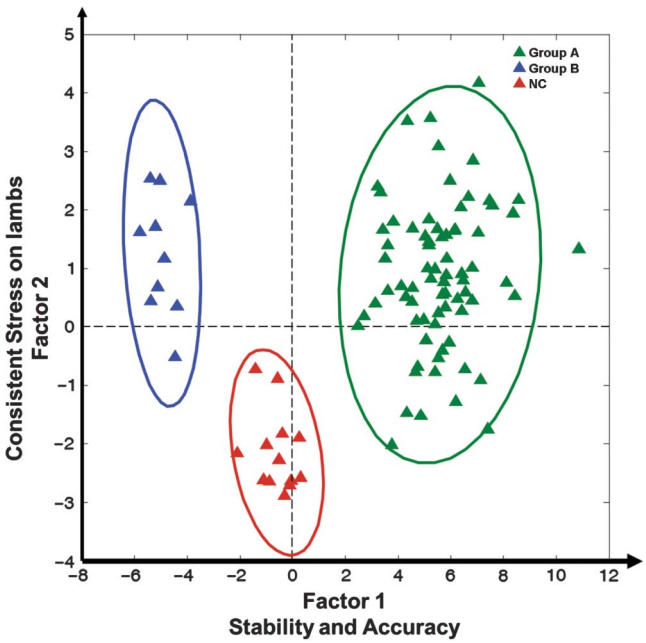
Not Classified

Thirteen participants in this study were NC (not members of either Group A or B). As depicted in Figure 4, participants in the NC group were evenly distributed along the linear factor that distinguished participants in Groups A and B. Comparison of this nonhomogeneous group with the other groups would not meet the required assumption of formal statistical analysis. Nonetheless, identification of distinguishing characteristics of this group of children was of interest. To identify distinguishing characteristics of participants in the NC group, an additional forward-stepping DA was run using the original 53 measures. An additional DA factor was necessary to distinguish this group of participants, as displayed in Figure 5. Participants in the NC group scored relatively higher on DA Factor 2 than the participants in Groups A and B. Consistent with higher scores on DA Factor 2, participants in the NC group had comparably higher scores on measures of acoustic variability on productions of both iambic and trochaic stress (Indices 8 and 9 in Supplemental Table 1) and lower average words per utterance (Index 51). Variability in lexical stress marking is consistent with findings associating unstable lexical stress with childhood apraxia of speech (Shriberg et al., 2003; Skinder, Connaghan, Strand, & Betz, 2000; Velleman & Shriberg, 1999), supporting a hypothesis that participants in the NC group may be a subgroup consistent with this diagnosis; additional data would be needed to support this hypothesis. Furthermore, these participants are not necessarily a cohesive subgroup, and their comparable performance on these tasks cannot be characterized as a subgroup of SSD without further analysis and validation, which was beyond the scope of this experiment.

Clinical Implications

SLPs, as with all health care providers, have increasingly been charged with adopting an evidence-based, patient-centered approach to clinical decision making. This enhanced level of clinical decision making requires the integration of patient preferences and research evidence with

Figure 5. Two-factor DA results to identify factors that maximally separated Group A, Group B, and the NC group. Positive values on DA Factor 1 (abscissa) were correlated with high performance on measures of movement stability (e.g., coefficient of variation on lower lip displacement) and phonemic accuracy. Positive values on DA Factor 2 (ordinate) were correlated with variable acoustic marking of lexical stress and lower scores on average words per utterance. Ellipses represent 99% confidence intervals.



clinical expertise. Traditionally, the clinician's experience and acumen have driven the identification of children with SSD who are suspected to have motor speech involvement, because the absence of physiological markers precludes objective measures (Strand et al., 2013). Using kinematic measures, the results of the current study provided empirical support and physiological signs for the existence of a motor speech subtype of SSD. Moreover, behavioral measures that exhibit the potential to be used as diagnostic markers for an MSD-NOS subclassification were identified. For example, the most salient behavioral marker for Group B was a child's difficulty imitating lexical stress. These children imitated bisyllables with trochaic stress with less than 50% accuracy and iambic stress with less than 5% accuracy. Children in Group A, consisting of children who did not evidence poorer performance on measures of motor control, imitated bisyllables with these stress patterns with 88% and 72% accuracy; these levels are similar to those of children with typical speech acquisition, who imitated these patterns with 82% and 73% accuracy (Vick et al., 2012). The SUBARP analysis identified children in Group B using a rule-based threshold of 17% or lower accuracy in the imitation of iambic targets.

Implementation of this algorithmic rule into use as a diagnostic marker might reasonably incorporate performance on imitation of bisyllables in a battery of tasks for

Downloaded From: http://jslhr.pubs.asha.org/ by a Health Sci Learning Ctr User on 04/14/2015
Terms of Use: http://pubs.asha.org/ss/Rights_and_Permissions.aspx

identifying children with motor speech disorders. Rounding to 20% for convenience, a screening task might require five iambic bisyllable imitations. Production of fewer than two accurate imitations would be noted as a risk factor in a child's potential classification as having deficits in speech motor control. The sensitivity for predicting Group B membership using this diagnostic marker for the sample in the current study (i.e., including Groups A and B and NC participants) was 100% and specificity was 94%, which suggests that bisyllabic iambic stress imitation could provide good resolution for identifying children with MSD-NOS. This simple imitative task could easily be incorporated into diagnostic protocols and is, in fact, part of the recently proposed Dynamic Evaluation of Motor Speech Skill (Strand et al., 2013). As part of an overall assessment of motor speech skill, the Dynamic Evaluation of Motor Speech Skill scores the first attempt of lexical stress imitation as prosodically correct or incorrect. The addition of five repetitions of a lexical stress bisyllable imitation might add to the validity and reliability of this measure, especially using a threshold of 20% accuracy.

Considering the high prevalence of SSD in preschool-age children (15.6%; Campbell et al., 2003) and the fact that most of these children will normalize by early elementary school, clinicians must also use evidence-based guidelines for presenting a child's likely prognosis in the support of planned services, including therapeutic duration and intensity. In addition to incorporating measures of speech motor control, such as the Dynamic Evaluation of Motor Speech Skill, into diagnostic practice, population prevalence estimates may guide health care policy and treatment guidelines regarding the burden on health care and development presented by SSD. Findings from the current study indicated that the estimated prevalence of preschool children with SSD without signs of motor involvement is 67.8%–84.8% (i.e., using 95% CI). The characteristics of these children were likened to the class of SSD termed SD in the SDCS. Children meeting SDCS criteria for SD are proposed to have individual and multiple causal pathways to their speech deficits, with data currently unavailable on normalization rates among the three putative subtypes of SD described in Shriberg et al. (2010a).

With this prevalence estimate, the remaining 15%–30% of preschool children with SSD would fall in the MSD subclass of the SDCS and would be at the highest risk for persistent SSD. This finding is consistent with the estimate that 25% of children with SSD will have speech features that persist past age 6 years (Flipsen, 2003; Goozée et al., 2007; Shriberg et al., 2010a). Direct measurement of speech movement in the present study indicated that 4%–16% of children with SSD exhibit deficits in speech motor control consistent with the characteristics identified in Group B. These deficits were consistent with those posited for the SDCS MSD-NOS subclass, a placeholder for children with evidence of a motoric impairment that is not consistent with childhood apraxia of speech or dysarthria (Shriberg et al., 2010a). Delaney and Kent (2004) estimated the prevalence of childhood apraxia of speech to be 3.4%–4.3% among children with SSD. The remaining proportion of children with SSD would then be accounted for by childhood dysarthria. These estimates should help guide health care policy and clinical decision making for preschool children with SSD.

Methodological Implications

The current findings support the utility of the SUBARP as a useful alternative to conventional subgrouping approaches, such as cluster analyses, for studies attempting to identify conceptually and clinically informative subgroups of a disorder using a data-driven approach. The SUBARP approach offered several advantages for this analysis, including the ability to identify small subgroups, using data-driven decision making to establish the validity of subgroups, and providing interpretable information about which measures distinguish each identified subgroup. This is in contrast to other machine-learning methods, such as support vector machines, in which the generated rules are mathematical formulas that are challenging to interpret. These key benefits of SUBARP made it ideal for the current study and demonstrated the feasibility of it and other subgroup discovery methods for comparable investigations in the behavioral and social sciences that seek to identify subgroups.

Limitations

Although the size and scope of this investigation provided support for the assertion that there is a small subgroup of children with SSD who exhibit differences in motor speech performance, some limitations should be noted. Standardized measures of language and speech are not available for the current participants. These scores would provide a perspective on how children in the two subgroups would perform clinically and how standardized measures of performance may vary in the two groups. Future work should catalog a number of standard diagnostic measures along with performance on the tasks used in this study to provide a more robust clinical picture of children in the two subgroups. As suggested previously, it is unlikely that currently available standardized measures of speech competence would be sufficiently sensitive to differences in speech motor control. It is also worth noting that although all participants in the current study passed a screening for receptive language, the presence of concomitant expressive language disorders was not explicitly ruled out. Children with specific language impairment have been shown to have measurable differences in speech motor control, especially in production of contrastive lexical stress (Goffman, 1999). From the spontaneous speech sample, average words per utterance (Supplemental Table 1, Index 51) was measured to estimate expressive language function and verbal productivity. Children in Group B, who had comparably poor ability to imitate iambic lexical stress, did not perform differently from children in Group A on this measure, suggesting that children in the two groups had similar lexical productivity. Differences in syntax and vocabulary were not assessed, however.

Downloaded From: http://jslhr.pubs.asha.org/ by a Health Sci Learning Ctr User on 04/14/2015
Terms of Use: http://pubs.asha.org/ss/Rights_and_Permissions.aspx

Conclusions

The goal of the present study of SSD was to seek empirical support for subgroups within the population of children with SSD. Using a data-driven, algorithmic approach, evidence emerged for two groups whose performance contrasted reliably on measures that suggested differences in speech motor control. Using the SDCS as an organizing framework within which the two subgroups might be described, the larger of the two emergent groups (76%) was thought to be consistent with the class of SSD termed SD, whereas the smaller group (10.3%) was thought to be consistent with the subgroup of SSD provisionally termed MSD-NOS. Given the relatively low estimated population prevalence of the MSD-NOS subgroup (as low as 4.3% in this study), it was essential to use a subgroup discovery method with the capacity to identify individuals with SSD who share the characteristics that defined this group. Future work may extend these findings to other samples of preschool-age children with SSD using subgroup discovery methods, with the goal of including more measures of linguistic and lexical stress performance that may help to identify individuals with other types of pediatric motor speech disorders, particularly to discriminate among those with childhood apraxia of speech, dysarthria, and MSD-NOS.

Acknowledgments

Funding for this project was provided by the National Institute on Deafness and Other Communication Disorders (Grant R01 DC00822; principal investigator, Christopher A. Moore) and the American Speech-Language-Hearing Foundation (New Century Scholars Doctoral Scholarship to Jennell C. Vick). We are especially grateful to the participants and their families, who graciously volunteered their time for this project. In addition, we gratefully acknowledge those whose work was critical for participant recruitment, data acquisition, data extraction, and computer programming: Tammy Nash, Jill Brady, Dayna Pitcairn, Denise Balason, Stacey Pavelko, Mitzi Kweder, Katherine Moreland, Lakshmi Venkatesh, Sharon Gretz, Kevin Reilly, Roger Steeve, Kathryn Connaghan, Yumi Sumida, Alyssa Mosely, Rossella Belli, Ettore Cavallaro, Jeanette Wu, Jenny Morus, Kelsey Moore, Mary Reeves, Nicholas Moon, Dennis Tang, Adam Politis, Andrea Kettler, Laura Worthen, Dara Cohen, Heather Mabie, Rebecca Mental, Michelle Foye, and Greg Lee.

References

Bartnikowski, S., Granberry, M., Mugan, J., & Truemper, K. (2006). Transformation of rational and set data to logic data. In E. Triantaphyllou & G. Felici (Eds.), *Data mining and knowledge discovery approaches based on rule induction techniques* (pp. 253–278). New York, NY: Springer.

Bauman-Wängler, J. A. (2004). *Articulatory and phonological impairments: A clinical focus* (2nd ed.). Boston, MA: Pearson/Allyn & Bacon.

Bernthal, J. E., & Bankson, N. W. (2003). *Articulation and phonological disorders* (5th ed.). Boston, MA: Pearson/Allyn & Bacon.

Campbell, T. F., Dollaghan, C., Janosky, J. E., & Adelson, P. D. (2007). A performance curve for assessing change in Percentage of Consonants Correct—Revised (PCC–R). *Journal of Speech, Language, and Hearing Research, 50,* 1110–1119.

Campbell, T. F., Dollaghan, C. A., Rockette, H. E., Paradise, J. L., Feldman, H. M., Shriberg, L. D., … Kurs-Lasky, M. (2003). Risk factors for speech delay of unknown origin in 3-year-old children. *Child Development, 74,* 346–357.

Connaghan, K. P., & Moore, C. A. (2013). Indirect estimates of jaw muscle tension in children with suspected hypertonia, children with suspected hypotonia, and matched controls. *Journal of Speech, Language, and Hearing Research, 56,* 123–136. doi:10.1044/1092-4388(2012/11-0161)

Connaghan, K. P., Moore, C. A., & Higashakawa, M. (2004). Respiratory kinematics during vocalization and nonspeech respiration in children from 9 to 48 months. *Journal of Speech, Language, and Hearing Research, 47,* 70–84.

Davis, B. L. (2005). Clinical diagnosis of developmental speech disorders. In A. G. Kamhi & K. E. Pollock (Eds.), *Phonological disorders in children* (pp. 3–21). Baltimore, MD: Brookes.

Delaney, A. L., & Kent, R. D. (2004, November). *Developmental profiles of children diagnosed with apraxia of speech*. Poster presented at the annual convention of the American Speech-Language-Hearing Association, Philadelphia, PA.

Dodd, B. (1995a). *Differential diagnosis and treatment of children with speech disorder*. San Diego, CA: Singular.

Dodd, B. (1995b). Procedure for classification of subgroups of speech disorder. In B. Dodd (Ed.), *Differential diagnosis and treatment of children with speech disorder* (pp. 49–64). San Diego, CA: Singular.

Dodd, B., & McCormack, P. (1995). A model of speech processing for differential diagnosis of phonological disorders. In B. Dodd (Ed.), *Differential diagnosis and treatment of children with speech disorder* (pp. 65–89). San Diego, CA: Singular.

Felici, G., Sun, F., & Truemper, K. (2006). Learning logic formulas and related error distributions. In E. Triantaphyllou & G. Felici (Eds.), *Data mining and knowledge discovery approaches based on rule induction techniques* (pp. 193–226). New York, NY: Springer.

Felici, G., & Truemper, K. (2002). A MINSAT approach for learning in logic domain. *INFORMS Journal of Computing, 14,* 20–36.

Felici, G., & Truemper, K. (2005). The Lsquare system for mining logic data. In J. Wang (Ed.), *Encyclopedia of data warehousing and mining* (pp. 693–697). Hershey, PA: Idea Group Reference.

Flipsen, P. (2003). Articulation rate and speech-sound normalization failure. *Journal of Speech, Language, and Hearing Research, 46,* 724–737.

Gibbon, F. E. (1999). Undifferentiated lingual gestures in children with articulation/phonological disorders. *Journal of Speech, Language, and Hearing Research, 42,* 382–397.

Gibbon, F. E., & Wood, S. E. (2002). Articulatory drift in the speech of children with articulation and phonological disorders. *Perceptual & Motor Skills, 95,* 295–307.

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. (2005). Assessment and classification: An integrative model of language and motor contributions to phonological development. In A. G. Kamhi & K. E. Pollock (Eds.), *Phonological disorders in children* (pp. 51–64). Baltimore, MD: Brookes.

Goozée, J., Murdoch, B., Ozanne, A., Cheng, Y., Hill, A., & Gibbon, F. (2007). Lingual kinematics and coordination in speech-disordered children exhibiting differentiated versus undifferentiated lingual gestures. *International Journal of Language & Communication Disorders, 42,* 703–724.

Downloaded From: http://jslhr.pubs.asha.org/ by a Health Sci Learning Ctr User on 04/14/2015
Terms of Use: http://pubs.asha.org/ss/Rights_and_Permissions.aspx

Green, J. R., Moore, C. A., Higashikawa, M., & Steeve, R. W. (2000). The physiologic development of speech motor control: Lip and jaw coordination. *Journal of Speech, Language, and Hearing Research, 43,* 239–255.

Green, J. R., Moore, C. A., & Reilly, K. J. (2002). The sequential development of jaw and lip control for speech. *Journal of Speech, Language, and Hearing Research, 45,* 66–79.

Green, J. R., Moore, C. A., Ruark, J. L., Rodda, P. R., Morvee, W. T., & VanWitzenburg, M. J. (1997). Development of chewing in children from 12 to 48 months: Longitudinal study of EMG patterns. *Journal of Neurophysiology, 77,* 2704–2716.

Green, J. R., & Wilson, E. M. (2006). Spontaneous facial motility in infancy: A 3D kinematic analysis. *Developmental Psychobiology, 48,* 16–28.

Herrera, F., Carmona, C. J., González, P., & del Jesus, M. J. (2011). An overview on subgroup discovery: Foundations and applications. *Knowledge and Information Systems, 29,* 495–525.

Lavrač, N., Cestnik, B., Gamberger, D., & Flach, P. (2004). Decision support through subgroup discovery: Three case studies and the lessons learned. *Machine Learning, 57,* 115–143.

McSweeny, J. L., & Shriberg, L. D. (1995). Segmental and suprasegmental transcription reliability (Phonology Project Tech. Rep. No. 2). Madison: Waisman Center, University of Wisconsin—Madison.

Moore, C. A., Caulfield, T. J., & Green, J. R. (2001). Relative kinematics of the rib cage and abdomen during speech and nonspeech behaviors of 15-month-old children. *Journal of Speech, Language, and Hearing Research, 44,* 80–94.

Moore, C. A., & Ruark, J. L. (1996). Does speech emerge from earlier appearing oral motor behaviors? *Journal of Speech, Language, and Hearing Research, 39,* 1034–1047.

Mugan, J., & Truemper, K. (2008). Discretization of rational data. In G. Felici & C. Vercellis (Eds.), Mathematical methods for knowledge discovery and data mining. *IGI Global-information science reference* (pp. 1–23). Hershey, PA: IGI Global.

Robbins, J., & Klee, T. (1987). Clinical assessment of oropharyngeal motor development in young children. *Journal of Speech and Hearing Disorders, 52,* 271–277.

Ruark, J. L., & Moore, C. A. (1997). Coordination of lip muscle activity by 2-year-old children during speech and nonspeech tasks. *Journal of Speech, Language, and Hearing Research, 40,* 1373–1385.

Shriberg, L. D., Allen, C. T., McSweeny, 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., McSweeny, J. L., & Wilson, D. L. (1997a). The percentage of consonants correct (PCC) metric: Extensions and reliability data. *Journal of Speech, Language, and Hearing Research, 40,* 708–722.

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

Shriberg, L. D., Ballard, K. J., Tomblin, J. B., Duffy, J. R., Odell, K. H., & Williams, C. A. (2006). Speech, prosody, and voice characteristics of a mother and daughter with a 7;13 translocation affecting FOXP2. *Journal of Speech, Language, and Hearing Research, 49,* 500–525.

Shriberg, L., Campbell, T. F., Karlsson, H. B., Brown, R. L., McSweeny, 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., Fourakis, M., Hall, S. D., Karlsson, H. B., Lohmeier, H. L., McSweeny, J. L., … Wilson, D. L. (2010a). Extensions to the Speech Disorders Classification System (SDCS). *Clinical Linguistics & Phonetics, 24,* 795–824. doi:10.3109/02699206.2010.503006

Shriberg, L., Fourakis, M., Hall, S. D., Karlsson, H. B., Lohmeier, H. L., McSweeny, J. L., … Wilson, D. L. (2010b). Perceptual and acoustic reliability estimates for the Speech Disorders Classification System (SDCS). *Clinical Linguistics & Phonetics, 24,* 825–846.

Shriberg, L. D., Lewis, B. A., Tomblin, J. B., McSweeny, J. L., Karlsson, H. B., & Scheer, A. R. (2005). Toward diagnostic and phenotype markers for genetically transmitted speech delay. *Journal of Speech, Language, and Hearing Research, 48,* 834–852.

Shriberg, L. D., & Lof, G. L. (1991). Reliability studies in broad and narrow phonetic transcription. *Clinical Linguistics & Phonetics, 5,* 225–279.

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., Lohmeier, H. L., Strand, E. A., & Jakielski, K. J. (2012). Encoding, memory, and transcoding deficits in childhood apraxia of speech. *Clinical Linguistics & Phonetics, 26,* 445–482. doi:10.3109/02699206.2012.655841

Shriberg, L. D., Potter, N. L., & Strand, E. A. (2011). Prevalence and phenotype of childhood apraxia of speech in youth with galactosemia. *Journal of Speech, Language, and Hearing Research, 54,* 487–519. doi:10.1044/1092-4388(2010/10-0068)

Shriberg, L. D., & Strand, E. A. (2014, February). *A diagnostic marker to discriminate childhood apraxia of speech from speech delay*. Paper presented at the 17th Biennial Conference on Motor Speech: Motor Speech Disorders & Speech Motor Control, Sarasota, FL.

Shriberg, L. D., Tomblin, J. B., & McSweeny, J. L. (1999). Prevalence of speech delay in 6-year-old children and comorbidity with language impairment. *Journal of Speech, Language, and Hearing Research, 42,* 1461–1481.

Skinder, A., Connaghan, K., Strand, K., & Betz, S. (2000). Acoustic correlates of perceived lexical stress errors in children with developmental apraxia of speech. *Journal of Medical Speech-Language Pathology, 8,* 279–284.

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.

Smith, A., & Zelaznik, H. N. (2004). Development of functional synergies for speech motor coordination in childhood and adolescence. *Developmental Psychobiology, 45,* 22–33.

Steeve, R. W., & Moore, C. A. (2009). Mandibular motor control during the early development of speech and nonspeech behaviors. *Journal of Speech, Language, and Hearing Research, 52,* 1530–1554. doi:10.1044/1092-4388(2009/08-0020)

Strand, E. A., McCauley, R. J., Weigand, S. D., Stoeckel, R. E., & Baas, B. S. (2013). A motor speech assessment for children with severe speech disorders: Reliability and validity evidence. *Journal of Speech, Language, and Hearing Research, 56,* 505–520.

Tomblin, J. B. (1989). Familial concentration of developmental language impairment. *Journal of Speech and Hearing Disorders, 54,* 287–295.

Truemper, K. (2009). Improved comprehensibility and reliability of explanations via restricted halfspace discretization. In

Downloaded From: http://jslhr.pubs.asha.org/ by a Health Sci Learning Ctr User on 04/14/2015
Terms of Use: http://pubs.asha.org/ss/Rights_and_Permissions.aspx

P. Perner (Ed.), *Machine learning and data mining in pattern recognition* (pp. 1–15). Berlin, Germany: Springer.

Velleman, S. L., & Shriberg, L. D. (1999). Metrical analysis of the speech of children with suspected developmental apraxia of speech. *Journal of Speech, Language, and Hearing Research, 42,* 1444–1460.

Vick, J. C., Campbell, T. F., Shriberg, L. D., Green, J. R., Abdi, H., Rusiewicz, H. L., … Moore, C. A. (2012). Distinct developmental profiles in typical speech acquisition. *Journal of Neurophysiology, 107,* 2885–2900.

Walsh, B., Smith, A., & Weber-Fox, C. (2006). Short-term plasticity in children's speech motor systems. *Developmental Psychobiology, 48,* 660–674.

Wiig, E. H., Secord, W., & Semel, E. (1992). *Clinical Evaluation of Language Fundamentals—Preschool*. San Antonio, TX: Psychological Corporation.

Wohlert, A. B., & Smith, A. (2002). Developmental change in variability of lip muscle activity during speech. *Journal of Speech, Language, and Hearing Research, 45,* 1077–1087.

Ziegler, W., Staiger, A., & Aichert, I. (2010). Apraxia of speech: What the deconstruction of phonetic plans tells us about the construction of articulation language. In B. Maassen & P. van Lieshout (Eds.), *Speech motor control: New developments in basic and applied research* (pp. 3–21). Oxford, England: Oxford University Press.

Appendix (p. 1 of 2)

Explanation of SUBARP

SUBARP is a subgroup discovery method, a technique falling under the umbrella of machine learning. It seeks to identify comprehensible groups within a set of data. The technique identifies patterns in the data that create rules that are intended for interpretation using the measures in the data set. In SUBARP, the data set is divided into training and testing sets of about equal size. In the terminology of subgroup discovery, measures of the data sets are called *attributes.* Thus, the values of the measures for an individual child are the values of the attributes of a record of the training or testing set. In an iterative process, the method declares one attribute to be a target and then tries to explain the variations in the values of that target by the values of the remaining attributes, where the explanations are humanly comprehensible rules. The method is designed to identify relatively rare subgroups within a small sample while processing a large number of measures. The technique also calculates statistical likelihoods that estimate the prevalence of any discovered subgroups within the population.

The algorithm processes each target separately in three steps.

1. Target values are discretized.
2. Values of the remaining attributes are used to explain the discretized target values.
3. From these explanations, interesting subsets of the records are derived.

Before commencing, the data set is divided equally into training and testing sets. SUBARP derives the important relationships and their significance from the training set. Then the testing set is used to determine whether these relationships exist in a second group. The probability of the coincidental discovery of groups with the same target and attribute rules is the measure of statistical significance. The result is a series of subgroups whose explanatory attributes achieve a level of significance that exceeds a predetermined threshold (e.g., 0.90).

Target Discretization

Let t be a target. The target t is discretized by the introduction of cut-points. Consider one such cut-point c. Let A be the subset of records with target value above the cut-point c and B be the subset of the remaining records. The cut-points are so chosen that A or B potentially contains an important subgroup. A more elaborate use of cut-points is also possible. There, two cut-points, c and d, say with $c < d$, are used, and the set A is the subset of records with target values falling into the interval defined by c and d. For large data sets, SUBARP uses an analysis of the pattern of target values to define the cut-points. For small data sets, such elaborate analysis is likely not useful. Instead, the target values are sorted and n cut-points are defined, where n ranges from 10 to 50. For each pair of subsets A and B, the remaining work of the algorithm is to explain the differences between these two subsets using the other attributes. The first step of that process involves feature selection.

Feature Selection

The set A is repeatedly partitioned into subsets A_1 and A_2; correspondingly, subsets B_1 and B_2 of B are specified. SUBARP finds a logic formula that achieves the value True on the records of A_1 and False on those of B_1. It then tests how often the same formula achieves True for A_2 and False for B_2. In all, 40 logic formulas are created. Within these 40 formulas, the frequency with which a given explanatory attribute is used suggests the importance of that attribute in explaining the differences between the sets A and B. A significance value is calculated for each explanatory attribute. Those with a significance value exceeding a threshold are selected for the next step, in which explanations of the differences between A and B are computed.

Downloaded From: http://jslhr.pubs.asha.org/ by a Health Sci Learning Ctr User on 04/14/2015
Terms of Use: http://pubs.asha.org/ss/Rights_and_Permissions.aspx

Appendix (p. 2 of 2)

Explanation of SUBARP

Computation of Explanations

With the selected attributes, two formulas are calculated. The first evaluates to True for the records of *A* and to False for those of *B.* The second formula evaluates to the opposite True–False values for the subsets. Both formulas consist of one or more clauses combined by *OR.* Each clause contains linear inequality terms combined by *AND.* For example, [($x < 4$) AND ($y > 3$)] OR [($x < 3$) AND ($y > 2$)]. Each of the two clauses in the example divided by *OR* is referred to as a factor. Both factors in the example contain specifications for critical values of the two attributes *x* and *y*. Because of the structure of the formulas, the following relationships hold: Each factor of the first formula evaluates to True for a subset of *A* and to False for the entire set *B*. That subset of *A* is a potentially important subgroup. By these definitions, the subgroup is completely specified by the target discretization condition defining *A* and the factor. Thus, it is given by some linear inequalities involving the target and the attributes occurring in the factor. Analogously, each factor of the second formula evaluates to True for a subset of *B* and to False for the entire set *A.* That subset of *B* is a potentially important subgroup, with a corresponding description involving linear inequalities. The next step selects factors, and thus subgroups, that are significant.

Factor Selection

The subgroups identified via the targets and factors may or may not characterize important configurations that are both interesting and useful. To estimate which case applies, SUBARP calculates a significance value for each subgroup, once more using the training data. For the discussion, consider the case where the subgroup is a subset of *A*. The significance value is the average of two values. The first value is the fraction of the size of the subgroup divided by the size of *A.* The second value is 1 minus the probability that a certain random process can generate the subgroup. That random process is called an *alternate random process* (ARP). It is one of several such ARPs used by SUBARP to evaluate whether a decision is possibly based on random effects or relies on structural results that very likely are not produced by some random process. Analogous computations involving the testing data instead of the training data produce a second significance value for each subgroup.

Evaluation of Subgroups

The average of the significance values obtained from the training and testing sets is assigned as overall significance. Only subgroups resulting from logic formulas with overall significance greater than 0.95 are considered potentially useful and subjected to the final test of statistical significance.

Test of Statistical Significance

Recall that each subgroup is defined by inequalities involving a target and the variables of a factor. Thus, there are target inequalities and factor inequalities. From the derivation of these inequalities, each record of the training set satisfying the factor inequalities also satisfies the target inequalities. If the subgroup is truly significant, then a similar result should hold for the testing records. That is, almost all testing records satisfying the factor inequalities should satisfy the target inequalities. This is tested via the binomial distribution using a suitably estimated probability that a randomly selected record satisfies the factor inequalities. The result of this test provides a statistical measure of the importance of the relationship, relative to chance. Only subgroups with significance values that are very small are considered for further interpretation.

Downloaded From: http://jslhr.pubs.asha.org/ by a Health Sci Learning Ctr User on 04/14/2015
Terms of Use: http://pubs.asha.org/ss/Rights_and_Permissions.aspx