

Distinct developmental profiles in typical speech acquisition

Jennell C. Vick,¹ Thomas F. Campbell,⁵ Lawrence D. Shriberg,³ Jordan R. Green,⁴ Hervé Abdi,⁵ Heather Leavy Rusiewicz,² Lakshmi Venkatesh,¹ and Christopher A. Moore¹

¹Department of Speech and Hearing Sciences, University of Washington, Seattle, Washington; ²Center for Research on Acquired and Neurologic Communication Disorders in Children, Children's Hospital of Pittsburgh, and The University of Pittsburgh, Pittsburgh, Pennsylvania; ³Waisman Center, University of Wisconsin-Madison, Madison, Wisconsin;

⁴Department of Special Education and Communication Disorders, University of Nebraska-Lincoln, Lincoln, Nebraska; and ⁵School of Behavioral and Brain Sciences, The University of Texas at Dallas, Richardson, Texas

Submitted 12 April 2010; accepted in final form 18 February 2012

Vick JC, Campbell TF, Shriberg LD, Green JR, Abdi H, Rusiewicz HL, Venkatesh L, Moore CA. Distinct developmental profiles in typical speech acquisition. *J Neurophysiol* 107: 2885–2900, 2012. First published February 22, 2012; doi:10.1152/jn.00337.2010.—Three- to five-year-old children produce speech that is characterized by a high level of variability within and across individuals. This variability, which is manifest in speech movements, acoustics, and overt behaviors, can be input to subgroup discovery methods to identify cohesive subgroups of speakers or to reveal distinct developmental pathways or profiles. This investigation characterized three distinct groups of typically developing children and provided normative benchmarks for speech development. These speech development profiles, identified among 63 typically developing preschool-aged speakers (ages 36–59 mo), were derived from the children's performance on multiple measures. These profiles were obtained by submitting to a *k*-means cluster analysis of 72 measures that composed three levels of speech analysis: behavioral (e.g., task accuracy, percentage of consonants correct), acoustic (e.g., syllable duration, syllable stress), and kinematic (e.g., variability of movements of the upper lip, lower lip, and jaw). Two of the discovered group profiles were distinguished by measures of variability but not by phonemic accuracy; the third group of children was characterized by their relatively low phonemic accuracy but not by an increase in measures of variability. Analyses revealed that of the original 72 measures, 8 key measures were sufficient to best distinguish the 3 profile groups.

child speech production; speech motor control; subgroup discovery; articulatory kinematics

TYPICAL DEVELOPMENT of speech motor control from 3 to 5 years of age is both promoted and constrained by a wide range of developmental factors, including linguistic, phonological, social, phonetic, phonemic, phonotactic (Gildersleeve-Neumann et al. 2008), physiological, cognitive, and communicative influences. These challenges in speech development are compounded by the rapid and nonuniform growth of the vocal tract and of craniofacial musculoskeletal structures (Fitch and Giedd 1999; Vorperian et al. 2005; Vorperian and Kent 2007). Although this complex developmental process exhibits a broad range of commonalities across children and across languages (e.g., early development of bilabials), the principle of motor equivalence and the essential variability of speech production would seem to make it unlikely that there is a singular motor solution across typically developing children to these influ-

ences and constraints. Nevertheless, it does seem likely that some solutions would provide advantages with respect to enhanced motor control or speech output such that the distribution of speech production variability may be nonuniform and nonmonotonic.

Despite the well-described development of speech through highly predictable milestones, typical mature speech production is characterized by ubiquitous variability (MacNeilage 1970) across observational domains within and across speakers. This striking paradox in motor development (i.e., a highly generative and variable motor behavior that develops universally with predictable characteristics and milestones) presents an opportunity to evaluate the interaction of widely divergent developmental influences in achieving a conceptually singular behavior: spoken communication. It may be that speech motor development is supported by the child's access to motor stereotypes (Thelen et al. 1987) that underlie oral motor and respiratory behaviors, which the child is able to access, fractionate, and remodel to achieve stabilized speech production (Green et al. 2002; Grillner 1981; Moore and Ruark 1996). Anatomic and physiological constraints and growth patterns may similarly drive patterns of early speech development. A well-ordered path to speech production might be expected from these regularities in the physiological characteristics of developing speech. Alternatively, children, like adults, may take advantage of the motor systems' capacity to achieve a singular target (i.e., a speech sound) through a continuously variable range of movement solutions (Hebb 1949).

The high variability inherent to speech development can be used to reveal potential developmental trajectories across levels of observation (e.g., acoustic, kinematic, behavioral) of oral motor behaviors in children. For example, there are instances within and across children for which variability can be expected to be relatively low (e.g., early-acquired speech sounds and words). In contrast, increased variability, within or across children, may reflect the instabilities entailed by coordinative development, craniofacial growth, or the demands of increasing lexical and phonetic complexity. Although each child's developmental context is unique such that the developmental profile exhibited will reflect high idiosyncratic variation, the ontogenetic and phylogenetic influences to develop speech may be so great that children are driven along a common pathway that reflects the trade-offs of motor capacities and behavioral (e.g., communicative, lexical) objectives. Identifying characteristic moments of changing coordinative stability that are common among subsets of children may reveal these

Address for reprint requests and other correspondence: J. C. Vick, Dept. of Psychological Sciences, Case Western Reserve Univ., Cleveland Hearing & Speech Center, 11635 Euclid Ave. - Rm. 333, Cleveland, OH 44106-7154 (e-mail: jennell@case.edu).

developmental pathways to mature speech motor control and further characterize children's growing foundation for speech production. It would also enhance our understanding of the constraints and goals in speech development.

The near universality of the development of speech motor control is evident in children's early productions of fluent, relatively intelligible speech, despite the absence of a number of later-appearing speech sounds. Acquisition of speech sounds follows a fairly predictable order, although the timing of acquisition varies (Shriberg and Kwiatkowski 1994). Moreover, when children produce words containing sounds that are absent from their repertoires, they generate errors that are predictable and consistent with their developmental stage, including common substitutions, omissions, and distortions of yet-to-be acquired speech sounds. These anticipated patterns of disruption reveal vulnerabilities in this developmental pathway and suggest the operation of potent motor stabilities (e.g., Thelen 1991), even in the presence of pervasive individual variability.

A range of physiological measures, including articulator displacement or muscle activation patterns, supports a developmental framework specific to speech. These measures have already provided insight into the emergence and development of speech production, revealing, for example, that early speech and vocal behaviors are distinct from other, earlier appearing, nonspeech oral motor behaviors (e.g., electromyographic patterns differentiate vocal and nonvocal behaviors, even in infants; Moore and Ruark 1996). Investigations of oral motor development have further revealed differentiation of speech and nonspeech behaviors over the course of several months in infants and young children (Green et al. 2000; Steeve and Moore 2009). The developmental trajectory of lip and jaw coordination during babbling and speech is nonlinear and exhibits differences in coordinative control across articulators through the first 6 yr of development (Green et al. 2000, 2002; Steeve and Moore 2009). These group descriptions do not, however, address the children's individual paths to mature speech motor control; these previous observations of relatively small groups of children precluded the discovery of distinct developmental paths or stages. The present investigation, using a much larger sample of children and measures, was designed to evaluate the range of individual pathways to speech motor development and to reveal distinct subgroups among these pathways.

Refinement of Variable Speech Motor Control

Children exhibit articulatory movements that are smaller, slower, and more variable than adults (Goffman and Smith 1999; Maner et al. 2000; Smith et al. 1995; Smith and Gartenberg 1984; Smith and Goffman 1998; Smith and Zelaznik 2004; Walsh and Smith 2002). This variability is consistently found for point measures of temporal (e.g., onset of lower lip depression) and spatial (e.g., maximum lower lip depression) measures of displacement (Goffman and Smith 1999; Maner et al. 2000; Smith et al. 1995; Smith and Gartenberg 1984) and for whole trajectory measures of stability (e.g., spatiotemporal index), which incorporate the entire movement gesture for a word or phrase (Goffman et al. 2008; Goffman and Smith 1999; Maner et al. 2000; Smith and Goffman 1998; Smith and Zelaznik 2004; Walsh and Smith 2002). This increased move-

ment variability in children compared with adults is exacerbated by increased utterance complexity, reflecting the effects of cognitive processing demands on motor control of speech movement (Maner et al. 2000; Smith and Zelaznik 2004; Walsh et al. 2006).

It is unclear how variability and instability in the coordinative infrastructure for speech are manifest in the child's speech output. It may be, for example, that the developing coordinative framework is most apparent only at the limits of the child's speech capabilities. The predictable segmental errors in children with speech sound disorders may reflect the rate-limiting effects of motor development (Campbell et al. 2003; Davis 2005), and more generally, motor development may be deterministic in the rate and sequence of speech development. These influences have not been studied thoroughly, however, since most studies of articulatory kinematics in children include only perceptually correct productions of target words or phrases. This omission leaves the full range of effects (e.g., cognitive, social, linguistic) unaddressed, since physiological constraints must be evident in erred as well as in correct productions. An important exception is the finding by Goffman et al. (2007) of few significant correlations between kinematic and segmental measures in typically developing children or in those with specific language impairment. This finding, that articulatory variability has no direct relationship to segmental errors, supports a model in which the developing speech communication system is decoupled from, and thus not entirely determined by, the physiological constraints of motor development. The present investigation was designed to further explore the relationships across levels of observation (i.e., behavioral, kinematic, and acoustic) in 3- to 5-yr-old children.

Experimental Objectives

The present investigation was designed to take advantage of the perspective on speech motor development afforded by children progressing through a very rapid and well-defined period of communicative growth. These observations were designed to isolate some of the behavioral elements of mastery for speech production (e.g., intelligibility, phonetic inventory) in conjunction with other key elements of coordinative development (e.g., acoustic or kinematic features). This investigation was also intended to reveal commonalities across children such that those who exhibit shared relationships across variables could be identified as following a common developmental path and that alternative developmental paths could also be distinguished. Accordingly, this experiment quantified and modeled a sample of physiological and behavioral characteristics of speech development in young children to identify different solutions to the development of speech motor control and to identify those measures that best differentiate among these subgroups of typical speech development. The long-term aim of this effort is to elucidate the most common means by which typically developing children manage the inherent variability afforded by their developing speech motor and cognitive systems so that typical variation can be distinguished from the approaches to speech motor control used by children with speech sound disorders.

The search for developmental pathways shared by multiple children (i.e., comprising a subgroup) was investigated using a multivariate approach designed to reveal distinct groupings,

stages, or continua in the measures obtained. The emergence of discrete statistical groups through the convergence of a broad array of empirical measures (i.e., through cluster analysis) required a relatively large number of children ($n = 63$) and a comparably large number of measures ($n = 72$) distributed across levels of observation (i.e., behavioral, acoustic, kinematic). Following cluster analysis, discriminant analysis was used to identify the smallest effective set of specific key variables that could uniquely discriminate and characterize the emergent subgroups of typical children. Sampling across a range of ages (3–5 yr) during which speech production is being refined allowed for a range of maturational profiles to emerge.

The preliminary descriptive statistics obtained also provided a comprehensive quantitative characterization of the physiological development of the labial, mandibular, and phonatory systems during speech and nonspeech behaviors in children with typical speech development aged 36–60 mo. This data set comprises a normative reference set for typical speech development during early childhood. As a practical consideration, incorporating a broad range of measures is critical to the statistical description of speech production in children of this age, but the eventual utility of this approach requires the reduction of this large preliminary array of measurements to those that contribute most to the variance in performance across and within subjects.

METHODS

Participants

Eighty-three participants between 36 and 57 mo of age with presumed typical speech development were initially enrolled in this study. Twenty participants were excluded: nine (10.7%) were found to have speech delay (i.e., according to the Speech Disorders Classification System); four (4.8%) dropped out of the study voluntarily; and the data from seven participants (8.4%) were not acquired due to technical problems that precluded further analyses (e.g., inadequate movement tracking). All protocols and analyses were completed for the remaining 63 participants. All children were monolingual speakers of English from the greater metropolitan area of Pittsburgh, PA. Inclusion criteria also included parental report of typical speech and language development with no history of treatment by a speech-language pathologist or audiologist. Exclusionary criteria included histories of developmental, neurological, or significant medical disorders. Written, informed consent was obtained from the parent or legal guardian of each participant. All protocols and procedures were approved and overseen by the Institutional Review Board at the University of Pittsburgh.

Table 1. *Categorical demographic variables*

Variable	Category	<i>n</i>	Percentage
Sex			
Male	1	27	42.9
Female	2	36	57.1
Family History			
Negative	0	58	92.1
Positive	1	5	7.9
Race			
Caucasian	1	51	81.0
African-American	2	9	14.2
Other	3	3	4.8

The 63 participants were characterized by the categorical demographic variables listed in Table 1. Continuous demographic variables (age, maternal age and education) are listed in *indexes 53–55* in Supplemental Table S1. (Supplemental data for this article is available online at the Journal of Neurophysiology website.) Participating children were between the ages of 36 and 57 mo with an average age of 46.1 mo. Fifty-seven percent of the participants were female (i.e., 36 of 63), 43% were male (i.e., 27 of 63), and 8% of the sample (5 of 63) reported a positive family history (at least 1 first-degree relative affected) of speech or language disorders. The age and years of education for each participant's mother (mean age 33.6 yr, range 21–45 yr; mean years of education 16.4; range 10–22 yr; see *indexes 54* and *55* in Supplemental Table S1) were also included in the statistical model following earlier predictors of disordered speech development by Campbell et al. (2003).

Inclusion Criteria

Participants exhibited speech development within normal limits, as determined from a 15-min conversational sample with a trained assistant. Narrow transcriptions of these speech samples were analyzed using the Programs to Examine Phonetic and Phonological Evaluation Records (PEPPER; Shriberg et al. 2001), which generated an error profile for each sample. These error profiles were compared with those of age-matched children in a lifespan reference database to confirm each child's inclusion with respect to typical speech development.

Each participant also exhibited receptive language skills within normal range (scaled score of 7 or above) as screened by the Linguistic Concepts subtest of the Clinical Evaluation of Language Fundamentals—Preschool (CELF-P; Wiig et al. 1992); oral structures within normal limits as evaluated by the Oral/Speech Motor Control Protocol (Robbins and Klee 1987); and normal hearing sensitivity, screened at 25 dB HL at 1, 2, and 4 kHz.

Experimental Tasks

The remaining experimental tasks were elicited during a second experimental session, which usually occurred within 2 wk of the first test session and included the targeted array of speech and nonspeech tasks. These tasks spanned a range of speech production complexity so that each child was expected to complete at least some of the tasks before reaching a level that exceeded his/her developmental capacities. Each type of task is represented in the data stream depicted in Fig. 1.

Speech sample (PEPPER). Each child engaged in a play session to elicit a 15-min conversational speech sample that was audio recorded for later transcription and analysis.

Lexical stress task. Target productions included five repetitions each of three trochaic (i.e., initial syllabic stress; e.g., BAbA) and three iambic (i.e., second syllable stress; e.g., baBA) bisyllables. The consonant varied among /b/, /p/, and /m/ (i.e., BAbA, baBA, PApA, paPA, MAmA, maMA), but the vowel was always /a/. Five imitations of each target were produced by the child in response to a recorded adult female model, yielding up to 30 imitations of contrastive lexical stress productions (i.e., 5 repetitions each of 6 targets) for each child. Production of distinct lexical stress contrasts have been shown

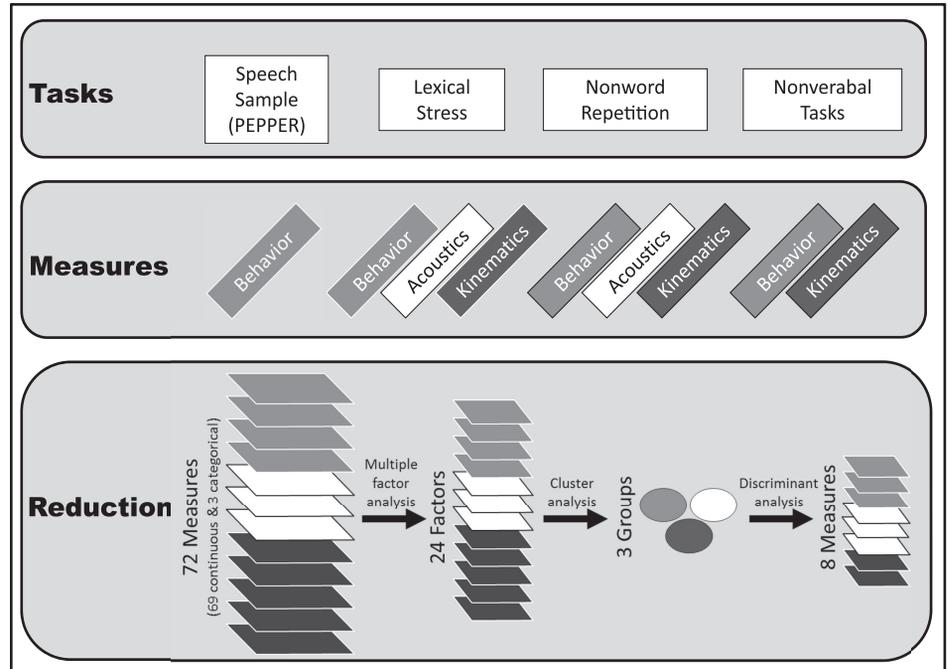


Fig. 1. Schematic of data reduction stream.

previously to reveal developmental differences among children in this age range (Shriberg et al. 2003). Occasionally, more or fewer than five repetitions of a target word were elicited from a child. In either case, all viable repetitions were included in the analysis, which yielded 1,840 productions across all speakers (i.e., 937 trochees, 903 iambs; see Table 2).

Nonword repetition task. Five repetitions of each of four nonword stimuli were elicited imitatively using a recorded female model. The four nonword stimuli were a subset of the nonwords included in the Syllable Repetition Task (SRT; Shriberg et al. 2009), were two and three syllables in length, and had primary stress on the first syllable (“bama,” “bada,” “bamana,” and “manaba”). These stimuli varied in difficulty and accordingly yielded an exaggerated range of variability; the systems and processes critical to repetition of nonwords include auditory perception, representation formation, memory

storage and retrieval, and motor planning (Shriberg et al. 2006). The range of difficulty presented by the increasing syllabic complexity of these stimuli, like the Lexical Stress Task, provided a means to titrate the production abilities of children in this age range. Occasionally, more or fewer than five repetitions of each target nonword were elicited from a participant; in either case, all repetitions from all participants were included, yielding a total of 1,362 productions. Details are shown in Table 2.

Nonverbal tasks. The nonverbal tasks included two behaviors, voluntary vertical jaw oscillations and chewing. The vertical jaw oscillations were produced imitatively in response to a live model (provided by the experimenter) elevating and depressing the mandible cyclically (i.e., opening and closing the mouth) through five cycles whenever possible. This task was repeated five times. The second task, chewing, was sam-

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

Type	Word	No. Attempted	Phonemics Correct		Stress Correct		Overall	
			<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Trochee	baba	325	284	87.4	321	98.9	233	71.7
	mama	308	285	92.5	292	94.7	237	76.8
	papa	304	286	94.1	284	93.4	237	78.1
	Total	937	855	91.2	897	95.7	707	75.4
Iamb	baba	307	259	84.4	303	98.8	209	68.1
	mama	295	274	92.9	232	78.5	183	62.0
	papa	301	267	88.7	266	88.4	201	66.9
	Total	903	801	88.7	801	88.7	593	65.7
Nonword	bada	342	309	90.4	n/a	n/a	291	85.0
	bama	357	306	85.7	n/a	n/a	304	85.1
	bamana	333	210	63.1	n/a	n/a	188	57.0
	manaba	330	173	52.4	n/a	n/a	160	48.5
	Total	1,362	998	73.3	n/a	n/a	943	69.2
	Grand total	3,202	2,654	82.9	1,697	82.9	2,243	70.0

Data includes number of tasks attempted and judged accuracy of phonemics and stress. “Overall” refers to the productions that were able to be used for all measures in the statistical model, which required correct phonemics and stress, as well as viable kinematics (i.e., without movement artifact). Totals are in bold type.

pled over two trials, during each of which the child chewed a single Goldfish cracker. Measures of performance on nonverbal tasks were included to distinguish speech-specific movement characteristics from those of more general orofacial behaviors.

Data Acquisition

Children were seated in a Rifton positioning chair fitted with a table during the experimental session. They were instructed to sit upright and to keep their hands on the table, holding a soft toy to avoid hand movements that might interfere with the acquisition of speech movement data.

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. In a few instances, 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 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 front view video recordings of the child's face. An infrared camera and light source (Burle TC351A) were used to illuminate small 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 forehead to provide landmarks for correction of head movement by motion tracking software. After each videotape was reviewed and logged, targeted task events were digitized for subsequent parsing and analysis. The sampling rate for the kinematic data was 60 Hz. These movement records 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 (e.g., baseline drift).

Two independent computer-based movement tracking systems were used to extract the position of each marker in the frontal plane (i.e., vertical and lateral positions) in Cartesian coordinates from the digitized video recordings. The first movement tracking system was version 6.05 of Motus (Peak Performance). The second system, which offered significantly improved processing speed, was DS-MTT version 2 by Hennis, a custom MATLAB routine created for movement tracking. Concordance across the two systems was >90% on an overlapped sample of 15% of the total data set.

Data Processing and Standardizing

Acoustic parsing. Each acceptable token was parsed using the audio signal from the video camera as a reference. The experimenter listened to the segment and roughly parsed the onset and offset of the production. 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.

Analyses specific to the vocalic portion of each production required extraction of these segments from the word produced.

The vowels of individual syllables were closely parsed and analyzed using custom algorithms in MATLAB. The vowel onset was defined as the first positive-going zero crossing in the waveform when the waveform became periodic (i.e., vocalic). The vowel offset was defined as the final negative-going zero crossing in the periodic signal associated with the vowel. Audio playback also assisted the experimenter when completing fine acoustic parsing.

Jaw, upper lip, and lower lip kinematic parsing. Kinematic signals were parsed with reference to the vertical displacement of the jaw marker. The first-order derivative of vertical jaw position (i.e., velocity) was superimposed on the position trace to parse the vertical jaw displacement onset and offset boundaries (i.e., using zero crossings, as described by Green et al. 2000). The last zero crossing appearing in the velocity trace before the jaw depression for the vowel operationally defined the onset of the opening movement; the first velocity zero crossing during jaw depression for the final syllable marked the offset of final jaw elevation (see Fig. 2). These onset and offset points were used to parse the displacement trajectories from the upper and lower lip.

Because the movement of the jaw contributes substantially to the displacement of the lower lip, the raw record of mandibular position included the movements of both 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 represented the net lower lip movement (Green et al. 2000).

Nonverbal task parsing. Analysis of the nonverbal tasks (i.e., chewing and voluntary jaw oscillation) was limited to jaw marker movement (i.e., no other markers were analyzed). Samples that included fewer than three cycles or exhibited movement artifact (e.g., obscured markers) were excluded from the analyses. First and last chewing cycles were removed from each chewing trial, as well. Because the analyses for these tasks evaluated cycle-to-cycle variability, each cycle was demarcated. Jaw elevation-depression-elevation (open-close) cycles were parsed algorithmically, demarcating each cycle by its peak elevation (identified by the associated zero velocity point). Because of the irregular displacement signal associated with molar contact during chewing, numerous velocity zero crossings occurred during some instances of jaw elevation. In these cases, the algorithm specified the latest occurring zero crossing as the cycle onset/offset boundary.

Conversational speech sample transcription and analysis. The examiners followed a standard SDCS protocol to obtain the speech samples; all data reduction was accomplished using the Phonology Project Laboratory Manual (unpublished). Narrow phonetic transcription of the continuous speech samples was completed by two experienced transcribers using consensus transcription procedures comprising a set of diacritical notations that classified subphonemic differences in articulatory place, manner, voicing, and duration (Shriberg et al. 1984). Calculation of point-to-point interjudge (86.7%) and intrajudge (91.8%) agreement for consonants was based on observations of 10 children who were part of a larger sample that included the children from the present study as well as children with speech delay of unknown origin. These procedures were described in a prior report (Shriberg et al. 2010).

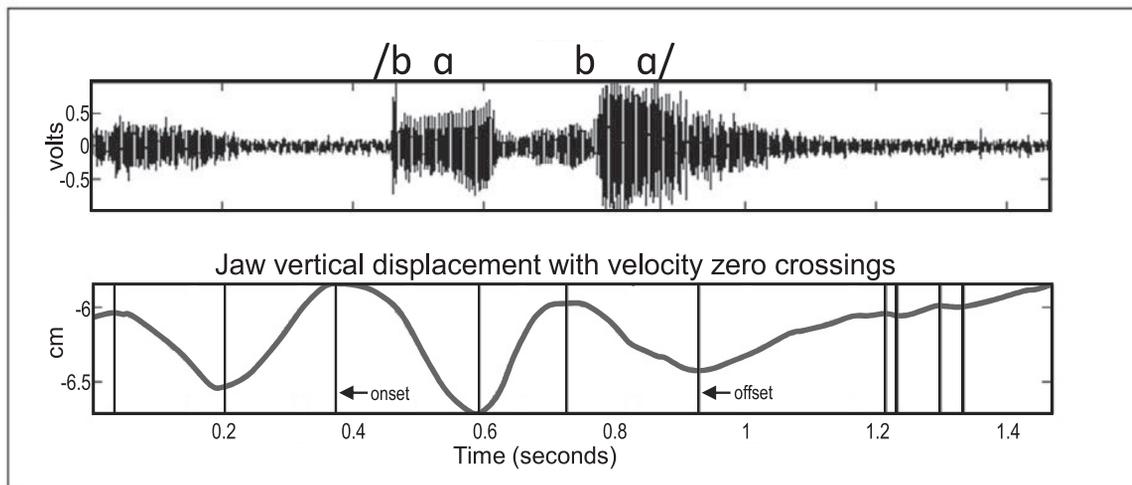


Fig. 2. Example of kinematic parsing using velocity zero crossings. *Top* graph plots the acoustic waveform for a child's production of "Baba." *Bottom* graph plots the vertical displacement of the mandibular marker. Vertical lines represent the velocity zero crossings. Onset and offset decisions are marked.

Perceptual analyses. Perceptual analyses were completed using the parsed audio signals, with separate perceptual judgment tasks for the lexical stress and the nonword repetition tasks. In the lexical stress perceptual task, audio files for each child's productions of each of the bisyllables were presented in randomized order including blocks of 10 children (i.e., about 300 audio files in each test block). Two listeners assessed whether the item was produced with the intended phonemic target and identified which syllable was stressed (first, second, or both, when even stress was perceived). All 1,840 contrastive stress productions were judged by two listeners. Joint probability concordance between the two listeners was 89.1%. Phonemic accuracy was similar between trochees (91.5%, 855/937; see Table 2) and iambs (88.7%, 801/903); however, imitative stress was markedly better for trochees (95.7%, 897/937) than for iambs (88.7%, 801/903).

The phonemic accuracy of the nonword repetition task productions was judged blindly (i.e., both to production target and participant) using an open-set, broad phonetic transcription task. The presented items were randomized by participant and production type. Three listeners judged each production. Transcription was finalized using a "best two of three" criterion [i.e., agreement by at least 2 of the 3 judges, which was reached on 95.2% (1,297/1,362) of the productions]. For the remaining 65 productions, the unedited video file was presented to enhance transcription task. Only when all syllables in a single token were produced accurately was the token scored as correct. By this criterion, 73.3% of the nonwords (998 of 1,362) were judged to be phonemically correct. Analysis by syllable length revealed decreasing accuracy from the first-attempted, two-syllable productions to the final, three-syllable productions in the experimental paradigm, with a drop in accuracy from the two-syllable (88.1%, 616/699 attempted) to the three-syllable (57.8%, 383/683) productions (bada: 90.4%,

309/342; bama: 85.7%, 306/357; bamana: 63.1%, 210/333; manaba: 52.4%, 173/330). Details are included in Table 2.

Construction and Derivation of the Modeled Data Set

Successful cluster analysis depends on adequate variances of the elements of a statistical model such that individual cases (participants) can converge into distinct and reproducible subpopulations. The present set comprised 69 continuous and 3 categorical variables spanning 3 orthogonal domains of observation: behavioral, acoustic, and kinematic. Multiple measures within each domain ensured that changes in performance within and across children would be observed (e.g., ceiling and floor effects were avoided by including observations across articulators of simpler and more difficult tasks, which appear earlier and later in motor development). In addition, although numerous measures have been investigated, it was not known which measures would be most sensitive to distinctions among emergent groups of preschool children with typical speech acquisition. A final benefit of this large set of measures is its future use as a benchmark data set for subsequent developmental studies of typical and disordered speech. Mean and range values for each measure (index) are reported in Supplemental Table S1.

Behavioral Measures

Proportion attempted (index 1). Although most participants attempted all of the verbal tasks, some children declined to imitate some of the presented tokens. Because the resultant empty data cells would result in these children self-selecting out of the analyses (perhaps for reasons unrelated to speech development) and contributing inappropriately and immeasurably to cluster identification, the overall proportion of tasks attempted by each participant was entered into the statistical model as an experimental measure. The overall mean for this

measure, 0.94 (ranging from 0.16 to 1.00), reveals that most children completed all of the tasks presented.

Proportion of productions with all phonemics correct (index 2). Perceptual analyses yielded the overall proportion of verbal productions transcribed that matched the intended target for each participant. Data from all attempted verbal productions (i.e., trochees, iambs, and nonwords) were collapsed into a single measure from each participant.

Proportion of each lexical stress type with correct stress (indexes 3 and 4). The proportions of attempted trochees and iambs produced with correct stress were included separately for each stress type and each child as input measures to the statistical model.

Proportion used (index 5). This value reports the proportion of all tokens produced by a child that were produced with both phonemic and lexical stress accuracy and had usable kinematic trajectories (i.e., free of movement artifact or markers obscured by hand motion). Children sometimes were minimally constrained and sometimes produced spurious movements out of frustration with a particular task, which reduced the number of usable tokens. Tokens produced with accurate phonemics and stress were included in acoustic analyses but were omitted from the kinematic analyses when necessary. Overall, 70% of the attempted verbal productions by the participants in the study were suitable for all levels of measurement.

PEPPER measures (indexes 62–69). The PEPPER measures (Shriberg et al. 2001) were used to generate error profiles for each child from a 15-min conversational speech sample. The PEPPER software environment was used 1) to classify participants' inclusion status as described previously and 2) to produce percentage scores on all of the competence variables (see below). It yielded a number of speech performance metrics that were included as input to the classification model. Each measure is intended to provide a metric of a unique speech domain within speech production:

PCC: percentage of consonants correct; a measure of phonetic accuracy, designed to quantify common and uncommon speech sound distortions, as systematized in the Appendix to Shriberg et al. (1997).

PCCR: percentage of consonants correct, revised; an index that does not count distortions (allophonic variation) as errors and provides a measure of phonemic accuracy. This measure provides greater distinction between children with normal speech acquisition and those with speech delay in lifespan reference data. It reflects the acquisition of the phonemes in the child's ambient community (Shriberg et al. 1997).

PVC: percentage of vowels correct.

PVCR: percentage of vowels correct, revised; parallel to PCCR, does not count distortions as errors.

PPC: percentage of phonemes correct.

PPCR: percentage of phonemes correct, revised; does not count distortions as errors.

II: intelligibility index; proportion of unintelligible words to total words produced, expressed as a percentage.

AWU: average words per utterance; an estimate of the child's expressive language function and verbal productivity, which varies greatly within typically speaking children in any given continuous speech sample. AWU is very highly correlated ($r > 0.90$) with mean length of utterance in preschool-aged children (Shriberg and Kwiatkowski 1994). This output measure reflects general lexical productivity, not grammar.

Acoustic Measures

Mean acoustic area ratio: trochees and iambs (indexes 6 and 7). For the lexical stress task, the primary acoustic analysis used was the Acoustic Area Ratio (adapted from Xie et al. 2004), which expresses the relative energy of the first syllable relative that of the second syllable. For each syllable, the mean amplitude was multiplied by the vowel duration to provide a measure of acoustic area (i.e., under the rectified and smoothed acoustic signal). Because the audio signals were not calibrated, amplitude was measured using the root mean square (RMS) of the signal in 20-ms windows. The mean amplitude for each participant and task was expressed as an average of these values. The Acoustic Area Ratio is the quotient of acoustic area of the first syllable and the second syllable; values >1.0 indicate that primary stress was on the first syllable, whereas those <1.0 indicate that primary stress was on the second syllable. The measure is similar to a metric used in the identification of subgroups of children with suspected apraxia of speech, the Lexical Stress Ratio (Shriberg et al. 2003). Overall, as expected, mean acoustic area ratio was greater for trochees (1.89) than for iambs (0.62). Individual data varied with ranges of 0.6–4.72 for trochees and 0.22–1.31 for iambs.

Coefficient of variation of acoustic area ratio: trochees and iambs (indexes 8 and 9). Variability in lexical stress production was evaluated using the coefficient of variation. This measure, the quotient of the standard deviation and the mean, was used to provide a normalized comparison of variability across widely divergent means and units of measure and was comparable for trochees (43.73) and iambs (40.81).

Word duration: two- and three-syllable nonwords (indexes 10 and 11). Word duration was measured for the two- and three-syllable nonwords from the burst of the /b/ or the initial glottal pulse of the /m/ to the final glottal pulse of the final /a/; glottal offset was operationally defined as the final negative-going zero crossing in the periodic signal associated with the vowel. Durations were collapsed within word length (i.e., across 2-syllable nonwords: "bama" and "bada," and across 3-syllable nonwords: "bamana" and "manaba"). Duration was substantially greater across children for three-syllable productions (1.45 s) than for the two-syllable productions (0.77 s).

Coefficient of variation of word duration: two- and three-syllable nonwords (indexes 12 and 13). These measures reflected the normalized variability of word duration for each participant. As shown in Table 2, the mean values for the group were comparable for the two- (11.78) and three-syllable nonwords (10.32).

Kinematic Measures

Mean maximum displacement: upper lip, lower lip, and jaw for both verbal (indexes 14–16) and nonverbal tasks (indexes 56 and 57). The maximum vertical displacement of the first syllable of all productions was measured for the upper lip, lower lip, and jaw. For nonverbal productions, the maximum displacement of the first cycle was measured for the jaw marker only. This measure was extracted automatically using custom scripts in MATLAB that identified the local minimum (jaw and lower lip) or maximum (upper lip) adjacent to the first vowel onset. Mean maximum displacement of the upper and lower lips produced by the group on verbal tasks were similar (0.22 and 0.24 cm, respectively), whereas the mean measure

for the jaw was much larger (0.72 cm). Mean values for jaw maximum displacement were slightly larger for nonverbal tasks, including voluntary jaw oscillations (1.08 cm) and chewing (1.00 cm).

Coefficient of variation for maximum displacement: upper lip, lower lip, and jaw for both verbal (indexes 17–19) and nonverbal tasks (indexes 57 and 58). The normalized variability for first-syllable (verbal tasks) and first-cycle (nonverbal tasks) maximum displacement was measured using the coefficient of variation. For the group, variability was similar for the upper and lower lips (39.3 and 38.0, respectively) and the jaw (31.2). Mean values for the jaw were similar for the nonverbal tasks, including voluntary jaw oscillation (38.0) and chewing (31.0).

Spatiotemporal index: trochees (indexes 20, 23, and 26), iambs (indexes 21, 24, and 27), three-syllable nonwords (indexes 22, 25, and 28), and nonverbal tasks (indexes 60 and 61) for upper lip, lower lip, and jaw (9 measures). Measures of movement stability were derived from the whole word movement records for trochaic, iambic, and three-syllable nonword productions for each single articulator. The spatiotemporal index (STI), developed by Smith et al. (1995), uses time and amplitude normalization of movement records for repeated whole words to reflect movement stability. Amplitude normalization of individual trajectories was accomplished by dividing each displacement record by its standard deviation. Linear temporal normalization involved interpolation of each amplitude normalized trajectory to 1,000 points using a cubic spline-fitting algorithm in MATLAB. The standard deviation was calculated across repetitions for each set of 20 consecutive points (i.e., 2% of the normalized sample per calculation); the sum of these 50 standard deviations yielded the STI.

For both the lexical stress task and the three-syllable nonword repetition task, STI was calculated for the upper lip, lower lip, and jaw markers' vertical displacements during each target production; only correct repetitions were included in these analyses. Overall, STI values for the upper lip and lower lip were higher than for the jaw, and values for the three-syllable nonwords were higher than for the two-syllable trochee and iamb productions.

A variant of the STI was used for the nonverbal tasks: the cyclic spatiotemporal index (cSTI), which quantifies the stability of repeated individual movement cycles (van Lieshout and Moussa 2000). Calculation of the cSTI used the parsed open-close cycles from the jaw vertical displacement records, which were time and amplitude normalized and aligned by start and end times. The sum of the 50 standard deviations calculated at every 20th sample constitutes the cSTI. This measure of variability was higher for chewing (26.52) than for voluntary jaw oscillations (15.19).

STI for lip aperture and jaw/lower lip aperture: trochees, iambs, and three-syllable nonwords (indexes 29–34; 3 measures). Two additional measures of combined articulatory movement stability were calculated for whole words. Analysis of the "net articulatory effect" arising from the movements of the upper and lower lips reflects the underlying articulatory goal and its associated functional synergies (Smith and Zelaznik 2004). The STI values of lip aperture trajectories (Kleinow and Smith 2006) were derived from the difference in trajectories for the upper and lower lip for each correct production. A similarly derived STI was calculated for the normalized difference tra-

jectories between the lower lip and the jaw displacements. Lip aperture STI values (indexes 29–31) were generally lower (reflecting lower variability) than lower lip/jaw STI values (indexes 32–34).

Convergence index: trochees, iambs, and three-syllable nonwords for upper lip, lower lip, and jaw (indexes 35–43, 9 measures). Following Goffman et al. (2007), a convergence index (CI) was calculated. This measure was derived from all attempted productions, including most notably those that contained phonemic and lexical stress errors, as well as those produced accurately. This index provided a measure of kinematic stability, even for children who produced few, if any, repetitions with phonemic or lexical stress accuracy.

CI/STI ratio: trochees, iambs, and three-syllable nonwords for upper lip, lower lip, and jaw (indexes 44–52, 9 measures). The ratio of the CI and STI was derived as a reflection of the relative stability of movement patterns across correct and incorrect productions. The CI/STI ratio has a value of 1.0 when only correct productions were obtained; higher ratios occurred when the variability was higher for the incorrect productions than the correct productions. This measure, for example, will reveal speakers whose variability increased dramatically for incorrect phonemic or lexical stress productions. For the group, the mean values for these ratios were all close to 1, although the ranges were large in some cases (e.g., 0.46–3.00 for upper lip on 3-syllable productions; index 46). Ratios <1 reveal that some repeated, incorrect productions were produced with less variability than correct productions for individual speakers.

Consideration of Age and Sex as Covariates

Because of well-described developmental and sex-related differences among many of these measures, it was essential to evaluate explicitly these potential effects. The linear relationship between age and each of the other 69 continuous measures was evaluated by calculating the Pearson correlation for each measure with age. Although all of the significant correlations were low using Cohen's criterion (i.e., $r < 0.5$), the highest linear relationship with age was noted in variables derived from the PEPPER conversational speech sample analysis. Specifically, the PCC, PCCR, PPC, and PPCR each had correlations >0.4 ($P \leq 0.001$). This result was highly anticipated, of course, because these age-normed measures were explicitly developed to be especially sensitive to speech development.

Sex differences were tested using a two-way between-groups analysis of variance (ANOVA); these potential effects were readily dismissed as deterministic for each of the continuous variables. There were no significant differences between males and females on any of the measures ($\alpha = 0.05$).

Data Reduction and Analysis

An obvious challenge associated with modeling a data set of 72 variables is narrowing consideration to those variables that are orthogonal, most representative of group behaviors, and sufficiently variable to inform interpretation of the model. Figure 2 provides a schematic of these data reduction steps and subsequent analyses.

Multiple factor analysis. Multiple factor analysis (MFA; Abdi and Valentin 2007; Escofier and Pagès 1990; Pagès and Husson 2005) is a variant of principal component analysis (PCA); it is used to integrate multiple data sets obtained from

a single set of observations. MFA was used in the present approach to reduce the dimensionality of the data preparatory to cluster analysis. Conceptually, MFA organizes variables into tables of similar variables, with each table matrix normalized before calculation of a global PCA. MFA is especially appropriate for applications like the present one in which the analysis of many different groups of variables are used to describe the same set of observations (Ding and He 2004; Gan et al. 2007).¹

Before MFA was performed, each variable was first centered so that its mean was zero, and then normalized so that the sum of its squared elements was equal to one. The MFA was performed with XLSTAT (version 2008; Addinsoft). The inclusion of factors in subsequent analyses was based on the cumulative proportion of the variance associated with each factor (Abdi and Williams 2010; Fielding 2007; Jolliffe et al. 1986). Subsequent analyses were based on the first 24 of 62 resulting factors, which accounted for 90.99% of the variance in the original data set. The remaining factors individually account for very small amounts of the total variance (i.e., 9% divided among the 38 remaining factors) and therefore could be dropped from further consideration.

Cluster analyses. The factors derived from the MFA were subjected to cluster analyses for post hoc identification of subgroups, using XLSTAT and two independent clustering algorithms. The first algorithm, hierarchical agglomerative clustering, was used to generate hypotheses about the number of clusters within the data; this step enhanced the subsequent cluster analysis approach, *k*-means, which requires the user to specify the number of clusters a priori. By using Euclidean distance to measure dissimilarity and Ward's method for agglomeration, the participant factor scores for each of the first 24 factors were submitted to a hierarchical agglomerative cluster analysis of all 63 children. The results of this clustering algorithm suggested that a six-cluster solution was optimal, although two of these clusters had only single cases. When these single cases were merged with adjacent clusters, the total number of groups was reduced to three, suggesting that a three-cluster solution was optimal.

Confirmatory clustering for this data set used *k*-means clustering, which is an iterative partitioning method of clustering that minimizes the within-cluster variation for a specified number (*k*) of clusters. The result is the creation of a family of clusters in which each entity belongs to just a single cluster (Fielding 2007; Mirkin 2005). This common clustering algo-

rithm (Gan et al. 2007) has been used effectively to classify children with specific language impairment (Conti-Ramsden et al. 1997) and children with speech delay of unknown origin (Venkatesh 2007).

To refine the clusters discovered using agglomerative hierarchical clustering, *k*-means clustering was run with two, three, four, five, six, and seven cluster solutions, using the groups found by the agglomerative clustering as initialization seeds. To validate and confirm the optimal cluster solution, external validity indexes were computed using the Cluster Validity Analysis Platform (CVAP, version 3.4; Wang 2007), a custom graphical-user interface in MATLAB.

Internal validity of a cluster solution uses only features inherent to the data set; indexes typically measure the cohesion of clusters and the separateness of the clusters in the solution. Two indexes were applied to the present solution, the Silhouette index and the Calinski-Harabasz index. The Silhouette index evaluated the overall quality of the cluster solution by measuring each cluster's tightness and separation. The optimality of the three-cluster solution was further supported by its highest average silhouette value. The Calinski-Harabasz index was used to confirm these results with a pseudo-*F* statistic, evaluating the cluster solution by evaluating the cohesion and separation of clusters (Calinski and Harabasz 1974). Again, the three-cluster solution was determined to be optimal.

Mean values for each of the 69 continuous raw measures for each of the three clusters are reported in Supplemental Table S1.

ANOVA. To identify significant differences in the continuous raw variables among clusters, a one-way, unbalanced ANOVA (XLSTAT) was used to evaluate each of the variables independently as a regressor on the three clusters, with application of a Bonferroni correction for pairwise comparisons. Pairwise comparisons are reported in Supplemental Table S1.

Descriptive discriminant analysis. A forward-stepping descriptive discriminant analysis was performed to determine what minimal combination of the original raw measures would best differentiate the three cluster groups. The nondifference of the three clusters' covariance values was confirmed with a Box test ($\chi^2 = 89.60$; $df = 182$; $P = 0.99$), which indicated that the data were in agreement with the assumptions of a discriminant model. The probabilities specified in the discriminant analysis were proportional to the cluster group sizes, because there was no a priori expectation of different probabilities for membership in a given cluster. The result was two discriminant functions (formed from linear combinations of 8 of the original 72 raw variables) that were found to maximize the differences between the three clusters of participants. Because stepwise methods can be associated with inflated chance associations and significance rates, jackknife resampling was used to cross-validate the discriminant function to decrease the probability of overestimating the intercluster distinctiveness (Tan et al. 2006). The hit rate on the cross-validated sample with resulting discriminant functions was 95.2%.

Figure 3 depicts the three clusters of individuals in the two-dimensional statistical space created by the two discriminant functions. The clusters' centroids appear as black stars, and the cases are surrounded by 95% confidence ellipses. The ellipses indicate the certainty of the positions of the clusters in the statistical space, with a probability of $1 - \alpha$ (95%) and

¹ Because of the large number of potentially interrelated measures in the present investigation, it was necessary to first transform the entire data set to reduce its dimensionality while maintaining as much of the variation as possible (Ding and He 2004; Gan et al. 2007). Typically, individual variables (vectors) are normalized before analysis, using vector normalization, such as *Z* score. The problem with such an approach in a very large data set with measures of diverse scope and variability is that blocks of measures with the greatest variance will dominate the outcome of the analysis. MFA mitigates this problem by normalizing all elements within the measure, thereby balancing the between-measure variance contributed to the analysis. Each group of similar variables, represented as a matrix, is normalized by dividing all elements by the first singular value of the matrix (i.e., the matrix equivalent of a standard deviation). This normalizing value for the measure matrix is calculated using a PCA: the square root of the first eigenvalue is the first singular value for the matrix. Once each table of measures is normalized, the matrices are then concatenated, and the resulting matrix is submitted to a grand PCA. The resulting principal components, or factors, linearly integrate the variables that contribute similarly to the variance of the data set (see Abdi and Valentin 2007 for a detailed presentation on this technique).

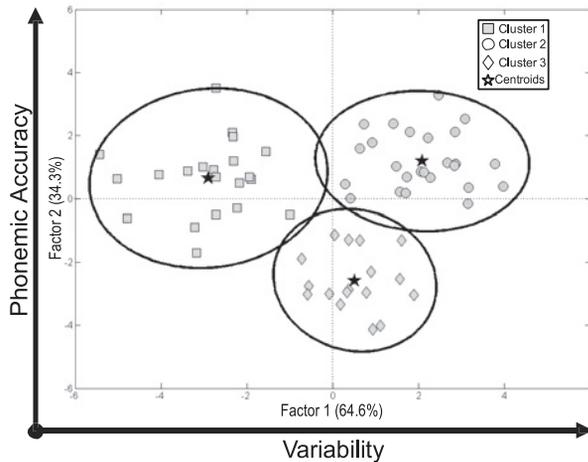


Fig. 3. Clusters in 2-dimensional discriminant function space with a star indicating each cluster's centroid. Cases for each cluster are encircled by 95% confidence interval ellipses. *Factor 1* (x-axis) represents variability; *factor 2* (y-axis) represents phonemic accuracy.

assuming a bivariate normal distribution. The minimal overlap of these ellipses indicates that these groups are distinct.

RESULTS

Inspection of the raw measures that contributed most substantially to the two dimensions (factors) of the discriminant space elucidated the most salient and distinct characteristics of the individuals comprising each of the three clusters. Pearson correlation statistics for the two factors and the eight most significant measures appear in Table 3. For each measure in Table 3, the correlation with *factors 1* and *2* are included (i.e., $F1 \rho$, $F2 \rho$) with the predominant correlations shown in bold. A high correlation value in Table 3 suggests that the measure accounted for a large proportion of the variance for that factor. In other words, *factors 1* and *2* can be described best using the measures with which they had the highest correlation values. In addition, these correlations can be used to describe the individuals who score at the extreme ends of *factors 1* and *2* in Fig. 3. For instance, individuals in *cluster 2* scored on the high end of *factor 1*, suggesting that they also score high on those measures that were positively correlated with *factor 1* and score low on those measures that were negatively correlated with *factor 1*.

Factor 1 (x-axis in Fig. 3) accounted for 64.6% of the variance in the data and is positively correlated with

- Acoustic area ratio of the first syllable to the second during trochaic productions (*index 6* in Supplemental Table S1),
- Variability of acoustic marking of stress in iambic productions (*index 19*),
- Jaw STI during iambic productions (*index 27*),
- Variability of jaw maximum displacement in the first syllable of all productions (*index 9*),
- Cycle-to-cycle variability of jaw movements during chewing (*index 61*), and
- Variability of lower lip movements during all productions that were meant to be trochees, regardless of accuracy (convergence index; *index 38*).

Participants with high scores for *factor 1* (i.e., *cluster 2*; see Fig. 3) exhibited greater variability across behaviors (verbal

and nonverbal) and across domains (kinematic and acoustic). This factor was labeled in Fig. 3 as “variability” to highlight the association of higher scores on *factor 1* with greater variability across a range of measures.

Factor 2 (y-axis in Fig. 3) accounted for 35.3% of the variance in the data and was positively correlated with

- PCCR (*index 63*), and
- Maximum displacement of the jaw during chewing (*index 57*).

Children with high scores on *factor 2* (i.e., *clusters 1* and *2*) had high scores on the PCCR and lowered their jaws more during chewing. This factor was labeled in Fig. 3 as “phonemic accuracy” to highlight the association of higher scores on *factor 2* with higher scores on measures of phonemic accuracy.

Subgroup Descriptions

The members of *cluster 1* ($n = 21$; 10 males and 11 females) generated measurements that were significantly different from those from members of either *cluster 2* or *cluster 3* on 40 different measures (see pairwise comparisons in Supplemental Table S1). Specifically relative to members of *cluster 2*, members of *cluster 1* produced all of the tasks, both verbal and nonverbal, with less variability (i.e., lower coefficients of variation for acoustic and kinematic metrics and lower spatio-temporal and convergence indexes: *indexes 8, 9, 17–20, 23–36, 38–41, 43, and 61*). These results were confirmed by the gathering of members of this group at the low end of *factor 1* in the discriminant analysis, which was associated with low measures of variability. In addition, compared with members of *cluster 2*, a greater proportion of the speech tasks produced by members of *cluster 1* were usable for kinematic analyses (i.e., more productions with accurate phonemics, lexical stress, and viable articulator tracking; *index 5*). Compared with members of *cluster 3*, members of *cluster 1* were significantly older (*cluster 1*: mean = 49.3 mo, SD = 5.6 mo; *cluster 3*: mean = 42.9 mo, SD = 5.0 mo; *index 53*); there was no significant difference in age between *clusters 1* and *2*. Members of *cluster 1* produced distinguishing speech and nonverbal tasks with less kinematic variability than members of the other clusters (e.g., *indexes 18, 29, and 41*). Accordingly, the *cluster 1* profile group was designated as “high stability.”

Members of *cluster 2* ($n = 25$; 6 males and 19 females) included a greater proportion of females (0.76) than members of *cluster 1* (0.52) or *cluster 3* (0.35). Members of *cluster 2*

Table 3. Correlations between the 8 variables identified by discriminant analysis and $F1 \rho$ and $F2 \rho$ from the discriminant analysis

Variable	$F1 \rho$	$F2 \rho$
Trochees: acoustic area ratio	0.404	0.120
Chewing: jaw maximum displacement	-0.087	0.500
Iambic: acoustic area ratio CV	0.475	0.226
PCCR	-0.279	0.669
Iambic: jaw STI	0.428	0.026
Jaw maximum displacement CV	0.508	0.026
Chewing: cSTI	0.532	-0.121
Trochees: lower lip CI	0.689	0.017

Values are Pearson correlation statistics for *factors 1* and *2* ($F1 \rho$ and $F2 \rho$) from discriminant analysis. Strong correlations are in bold type. CI, convergence index; cSTI, cyclic spatiotemporal index; CV, coefficient of variation; $F1$ and $F2$, factors 1 and 2 from discriminant analysis; PCCR, percentage of consonants correct, revised; STI, spatiotemporal index.

were more variable in their acoustic realizations of iambic stress tokens compared with members of the other clusters (*index 9*). Both point and dynamic measures of articulatory variability were significantly greater for members of this group compared with the others (*indexes 18, 26, 29, 31, and 41*). Accordingly, the *cluster 2* profile group was designated as “high variability.” Members of this cluster also distinguished themselves from members of the high stability profile group on the two measures of acoustic area ratio (*indexes 6 and 7*). For both trochaic and iambic productions, values for the acoustic area ratio were significantly higher, suggesting that first syllables were produced with greater emphasis (i.e., louder and longer) by the high variability group compared with those produced by members of the high stability group.

Members of *cluster 3* ($n = 17$; 11 males and 6 females) were significantly younger (42.9 mo) than members of the high stability profile group (49.3 mo), but no significant age difference was found when they were compared with members of the high variability group (45.6 mo; *index 53*). Their low scores on the phonemic accuracy dimension of the discriminant analysis (see Fig. 3) were consistent with the pairwise comparisons of the individual raw measures. Specifically, members of *cluster 3* scored lower on many of the PEPPER output variables (i.e., PCC, PCCR, PPC, PPCR; *indexes 62–67*) compared with members of the other two profile groups. For this reason, this profile group was designated as “low phonemics.” In addition,

members of the low phonemics profile group had jaw and lower lip maximum displacements that were significantly lower than those produced by members of the other two groups when chewing (*index 57*) and lower than those of the high stability profile group on verbal tasks (*indexes 15 and 16*).

Figure 4 presents standardized performance of the three profile groups relative to the whole sample mean on each of the eight measures that were found to best discriminate the groups. Significant pairwise comparisons from the ANOVA are indicated. The three profile groups were discriminated across levels of observation including behavioral (i.e., PCCR), phonatory (i.e., trochaic and iambic acoustic measures), and articulatory (i.e., verbal and nonverbal measures of displacement and variability).

In summary, characteristics of members of each of the three groups identified by cluster analyses were identified using pairwise post hoc ANOVA comparisons and placement in the statistical space created by discriminant analysis. The high stability profile group included 21 members whose performance was characterized by low levels of acoustic and kinematic variability across both verbal and nonverbal tasks and high phonemic accuracy. In contrast, the 25 members of the high variability profile group produced comparatively greater variability across acoustic and kinematic measures but had similar levels of phonemic accuracy to the high stability profile group. The 17 members of the low phonemics group scored

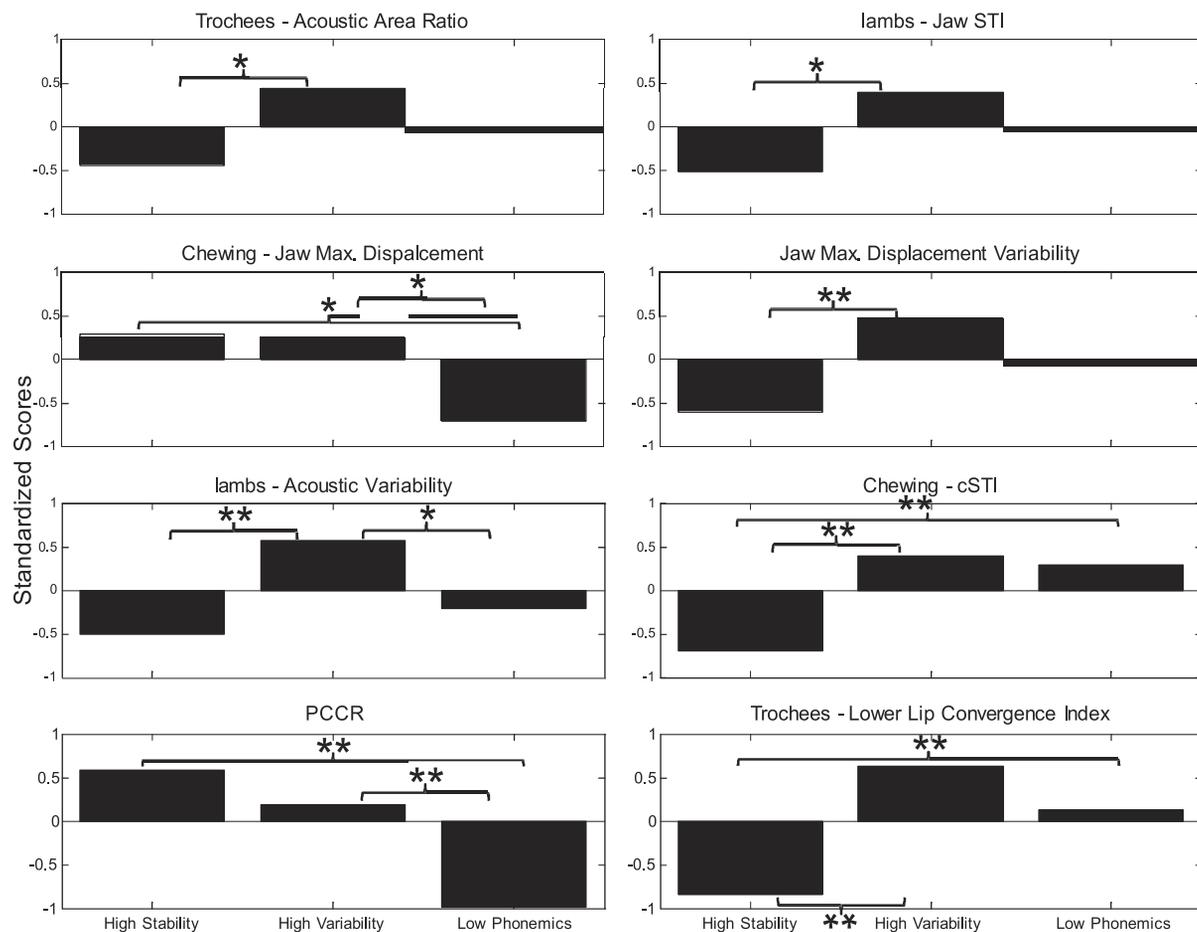


Fig. 4. Standard scores of each group (high stability, high variability, low phonemics) on each of the 8 variables found to maximally distinguish the groups. Significant pairwise comparisons are indicated: * $P < 0.017$; ** $P < 0.001$. cSTI, cyclic spatiotemporal index; Max., maximum; PCCR, percentage of consonants correct, revised; STI, spatiotemporal index.

comparatively lower on measures of phonemic accuracy, such as the PCCR and other output measures from the PEPPER analysis. Although these scores were significantly lower than the scores of the other two profile groups, they were still within the range of children with normal speech acquisition. Overall, significant pairwise comparisons between profile groups existed on 41 of the 72 raw measures. Results from a forward-stepping discriminant analysis suggested eight of the measures could be used to best discriminate the three profile groups. The eight measures included PCCR, five measures of acoustic and kinematic variability, an acoustic measure of lexical stress production, and a measure of jaw maximum displacement during the chewing task.

DISCUSSION

Analyses of this large sample of 3- to 5-year-olds' speech revealed three distinct developmental clusters. This finding alone, that the processes of speech development are nonuniformly distributed within and across children, weakens support for a model of speech production characterized by gradual, monotonic refinement and stabilization of early-appearing movement patterns and behaviors. Rather, the present findings support the alternate hypothesis that the many potent influences driving speech development, within and across children, elicit a range of behavioral characteristics, and that the speech production solutions available to young children are manifest in distinct kinematic/acoustic/behavioral profiles.

Beyond the identification and description of emergent clusters among these typically developing children, the present approach identified those measures that most specifically and accurately characterized each group. The initial modeling results revealed that the most distinctive descriptors included a combination of measures related to phonemic accuracy and articulatory/phonatory stability. Forward-stepping descriptive discriminant analysis was used to identify the 8 most discriminating measures among the original 72. These eight distinguishing variables are shown in Fig. 4. Of these eight, four were related to articulator movement variability: 1) the coefficient of variation of the maximum displacement for the jaw, 2) the spatiotemporal index of the jaw movements for iambic productions, 3) the convergence index of the lower lip movements for trochaic productions (like the STI, but including productions in error), and 4) the cyclic variability of the movement of the jaw during chewing. The fifth discriminant measure was 5) phonemic accuracy (PCCR), which enhances the ecological validity of these measures as sensitive to differences in speech output. The sixth and seventh measures were derived from speech acoustics: 6) the acoustic area ratio for production of trochees, and 7) the coefficient of variation of the acoustic area ratio for iambic productions. Finally, 8) the maximum displacement of the jaw during chewing distinguished the low phonemics group from the other two groups, which was consistent with several other measures of maximum displacement for the low phonemics group (i.e., during word production, jaw and lower lip maximum displacements were significantly smaller in this group than in the high stability group; upper lip maximum displacements were significantly smaller than those of the high variability group). It may be that the smaller jaw maximum displacements observed in this cluster may simply reflect morphological differences associ-

ated with this group being the youngest overall of the three identified groups. Maximal mandibular displacement has been found to be correlated with age and stature in children (Landtwin 1978) and with facial morphology in adults (Fukui et al. 2002), although the size of orofacial structures has been found to have little relationship with the scale of speech movements (Riely and Smith 2003). It is likely that these kinematic differences are a distinguishing characteristic of individuals in this group.

The three emergent speech profiles may have resulted from each group of children's differential weighting of widely divergent, yet powerful, influences that are in effect during speech development (e.g., communicative intent, craniofacial morphology, level of language or motor development, phonetic repertoire). These three identified clusters were broadly characterized by their 1) relatively high production stability, 2) relatively high production variability, or 3) relatively low phonemic accuracy. The statistical support and ecological validity of these profiles were strengthened by a modeling approach (MFA) and cluster analyses designed to maintain the interpretability of the findings throughout the process. The group characteristics relied directly on the surface measures (e.g., acoustic area ratios) so that each profile could be distinguished by established speech descriptors. The high stability group was characterized by low values for variability measures across both acoustic and kinematic domains and by high performance measures of phonemic accuracy (e.g., PCC); the high variability group was characterized by significantly higher variability on both acoustic and kinematic variability measures and with performance measures of phonemic accuracy that were similarly high as the high stability group; and the low phonemics group was characterized by low performance on measures of phonemic accuracy. Each of these clusters represented a distinct behavioral response or profile for the speech production tasks by each child.

During early speech motor development, children expand and refine their phonetic, phonemic, and lexical repertoires while experiencing rapid neurological, behavioral, and morphological growth. Prior studies have indicated, not unexpectedly, that these tumultuous conditions yield a high degree of heterogeneity in observed speech motor control and resulting phonetic behaviors (Goffman 1999; Holm et al. 2007; Smith and Gartenberg 1984; Smith and Goffman 1998; Smith and Zelaznik 2004). The present results reveal that, within these broad ranges of variation, it is possible to identify distinct subgroups within children, which suggests the possibility of alternative pathways to mature speech production. These distinct solutions to the control problems presented by speech development may reflect discrete individual responses to necessary trade-offs in early speech production, including such factors as articulatory accuracy versus phonotactic complexity, or word familiarity/mastery/frequency versus lexical expansion and word novelty. Children might be expected to react to these competing influences idiosyncratically, although the range of potential solutions appears to be limited to those exhibited by the three identified clusters. Adoption of specific speech production styles in response to specific goals (e.g., opting for high movement stability over phonotactic complexity) may lead to parallel developmental pathways, manifest in the present experiment as distinct clusters, such that the clusters identified comprise distinct groups of children who

have taken separate pathways to the goal of mature speech production.

Another possible interpretation of these findings of distinct developmental clusters is that each child passes sequentially through each or some of the identified groups, perhaps even repeatedly cycling through these clusters, during speech development. The present observations are consistent with the suggestion that stages of relative stability in developing speech motor control are interleaved with transitional periods of relative instability. It seems likely that speech development entails periods of relatively high or low movement stability and of relatively high or low phonemic accuracy. The present sample, which revealed children in each of these states, comprised a snapshot of each child's development and leaves open the alternative hypothesis that observing their productions at another time or under different conditions would reveal his or her membership in one of the other two groups. It seems unlikely that such shifts could happen very quickly, however, for two compelling reasons: 1) task effects were not observed in the present data set, and 2) it seems especially unlikely that a child's PCCR scores would exhibit the levels of volatility needed for a child to move abruptly from membership in one cluster to another. Thus these clusters appear to have emerged from developmental differences, rather than from task-responsive effects. Frequent longitudinal sampling would be necessary to test this suggestion empirically.

A sequential pathway involving these three profiles might be envisioned by proposing that each cluster represents a distinct period in the developing coordinative infrastructure of the speech production systems (e.g., increasing accuracy in predictive mapping of vocal tract shapes and movements to speech acoustics). Performance by members in the low phonemics profile group, for example, may reflect a state in which the child generates relatively stable speech production at the expense of a limited phonetic repertoire. Developmental influences promoting an expansion of this phonetic repertoire would promote lexical expansion and enhanced intelligibility but would likely come at the cost of transitioning through a state of relative instability, like that measured in the high variability profile group. Mastery of this expanded speech inventory would be expected to give rise to a high stability profile, until developmental influences once again promote transitioning through another period of instability to a higher stage of speech development. Even though age effects did not determine cluster membership, this progression is supported by one age-related post hoc result: children in the low phonemics group (mean age: 42.9 mo) were significantly younger than children in the high stability group (mean age: 49.3 mo), although the ages of children in the high variability group (mean age: 45.6 mo) were not statistically different from those in either of the other groups. It is critical to incorporate the likely effects of oral motor development into this scenario, however; differences in motor variability among the three profile groups were present across both verbal and nonverbal tasks, which suggested that these varying motor coordinative stabilities were not specific to the speech production system.

In any case, it cannot be determined from the present results whether membership in a cluster occurs in parallel across children (i.e., each child expresses the speech characteristics of a single cluster throughout development) or sequentially within each child (i.e., that each child will manifest the characteristics

of each cluster in order through different periods of development), or whether developmental periods elicit the characteristics of a specific cluster.

It is also important to note that although measures of phonemic accuracy were correlated positively with age, specific tests of this relationship revealed no group effect for age and any of the remaining measures. Similarly, there was no significant relationship observed between sex and any of the reported measures. These findings make it very unlikely that the observed clusters were driven by chronological age or sex effects. By contrast, Smith and Zelaznik (2004) found significant differences in phrase-level kinematic variability between 4- and 5-year-olds, as well as between males and females (males were more variable). A potential reason for this difference in findings was the result of the level of analysis: single nonwords were analyzed in the present study compared with phrases in the study by Smith and Zelaznik.

The present design also provided a description of the relationships of observations across speech production domains, generating a broad characterization of the physiological, behavioral, and acoustic dimensions of speech motor development of young children, and provided a test set for post hoc questions of speech development. For example, these data facilitated an evaluation of the hypothesis that children who exhibited the highest variability in production measures would produce the greatest number of phonemic errors. The CI/STI ratios (*indexes 44–52* in Supplemental Table S1; the CI and STI for each articulator and target production) compare variability of incorrect and correct productions with those of correct productions only and were used to evaluate this question. No statistically significant finding was obtained for this test, although it exemplifies the wide range of questions that remain to be addressed in this data set.

Broader Implications

The present results have raised several immediate questions, have important implications for the design of future studies of speech production of children at these ages, and also inform our interpretation of past findings. Research on speech development has most often relied on small sample sizes using measures in a single domain (e.g., speech kinematics) derived from a single subsystem (e.g., mandible). These prior investigations have framed our understanding of typical development, disordered speech production, and clinical practice. For example, it is widely understood that young children produce speech that is more variable than that of older children and adults (cf. Chermack and Schneiderman 1986; Eguchi and Hirsh 1969; Kent 1976; Kent and Forner 1980; Smith 1978; Tingley and Allen 1975). The present results narrowed these findings to specific types of variability for these target utterances.

A study of short phrase production by eight 4- and eight 7-year-old children by Smith and Goffman (1998), for example, reported mean lower lip plus jaw STI values of 24.1 for the 4-year-olds, which was significantly lower than the STI value of 18.5 obtained for the 7-year-olds. The present results included comparable measures for the STI of the jaw, but during trochaic bisyllables: children in the high variability group had mean STI values of 24.7, children in the high stability group had mean STI values of 18.6, and those in the low phonemics group had mean STI values of 20.7. The results of the present

high variability group replicate those reported by Smith and Goffman for 4-year-olds, although the high stability group exhibited STI values that were significantly lower than those of the high variability group. Despite the difference in context (i.e., nonword bisyllables in the present study vs. phrases in the Smith and Goffman study), the present results may provide a refined context for interpretation of these findings, demonstrating that the ubiquitous variability of speech production can be understood in a more detailed way, with some children adopting a speech production style that reduces this variability.

The present investigation has also demonstrated the feasibility and necessity of experimental designs that use samples that are sufficiently large to differentiate real, but subtle, differences in speech production. Moreover, multivariate analysis and clustering techniques were shown to provide interpretable results that will provide a more detailed understanding of speech production, developmental pathways, and disorders. Nevertheless, the current data set, although large by current research practices in developmental speech physiology, is still too small to provide a high-resolution description of the various developmental trajectories that result in mature speech production. For example, the full distribution and shape of the variations in developmental profiles could only be understood with a very large developmental sample that includes frequent longitudinal sampling.

Limitations

In addition to the limits posed by the cross-sectional design of the present investigation, these findings were also affected by empty cells in the data set. Lost data sometimes resulted from the failure of the system to track the optical targets consistently; automated identification of the flat reflectors was easily disrupted by extraneous movements. It is also worth noting that movement of the mandible was transduced and projected only in the coronal plane, which neglected other degrees of freedom, particularly rotation around the horizontal axis and anterior-posterior translation (Westbury 1988). Multiple-camera movement tracking systems are generally sufficient to overcome these limitations. An additional source of error is that associated with the use of a flesh point (e.g., the midline chin placement of the jaw marker) as an indicator of skeletal movement (Green et al. 2007); elimination of this error is challenging in young children, of course.

Conclusions

The primary aim of this study was to evaluate the speech production profiles of young children by using a large battery of measures from physiological, acoustic, and behavioral domains. Of the 72 measures obtained, 8 were found to be sufficient to distinguish 3 clusters of children. These findings suggest that the high variability that is typical of speech production is not uniformly observed across children; different groups of children exhibit significantly different degrees of kinematic, acoustic, and behavioral variability. This finding supports the suggestion that developing speakers employ distinct developmental profiles in producing speech. The present data set was not designed to determine whether these differences are exclusive across children (i.e., with different groups of children developing along parallel paths to achieve mature speech production) or within children (i.e., with each child

progressing along a pathway with separable, distinct periods). It remains for future investigations to resolve this question by sampling these behaviors longitudinally with relatively frequent observations.

Prior studies have established that variability characterizes measures of articulatory kinematics and acoustics at this period of development. The multivariate statistical techniques employed in this study enhanced the resolution of our understanding of speech motor development at these ages. High variability was observed in one of the identified groups of children, although the phonemic accuracy of the children in this group was not significantly different from that observed in the group of children with the lowest level of articulatory and acoustic variability. A third group was identified whose articulatory and acoustic variability was no different from that of the other two groups but whose phonemic accuracy was significantly lower.

These results have also further elaborated the interactions of speech production systems across levels of analysis, providing a better understanding of how observations at one level (i.e., labiomandibular movement) do not reliably predict observations at another level of analysis (i.e., phonemic productions). In the present sample, children in the high variability group produced speech sounds that were recognized as acceptable tokens of the target sounds, yet were quantitatively different from those produced by children in the high stability group.

ACKNOWLEDGMENTS

We are especially grateful to all of the participants and their families who graciously volunteered time for this project. In addition, we acknowledge the work of the individuals 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, 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, and Greg Lee.

Present address of C. A. Moore: Department of Veterans Affairs, Washington, DC.

Present address of L. Venkatesh: Sweekar Rehabilitation Institute for Handicapped, Secunderabad, Andhra Pradesh, India.

Present address of H. L. Rusiewicz: Department of Speech-Language Pathology, Duquesne University, Pittsburgh, PA.

GRANTS

Funding for this project was provided by National Institute on Deafness and Other Communication Disorders Grant R01 DC00822 (to C. A. Moore) and a American Speech Language Hearing Foundation New Century Scholars Doctoral Scholarship (to J. C. Vick).

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the authors.

AUTHOR CONTRIBUTIONS

Author contributions: J.C.V., T.F.C., L.D.S., J.R.G., and C.A.M. conception and design of research; J.C.V., T.F.C., L.D.S., and L.V. analyzed data; J.C.V., T.F.C., L.D.S., J.R.G., H.A., and C.A.M. interpreted results of experiments; J.C.V. prepared figures; J.C.V. drafted manuscript; J.C.V., T.F.C., L.D.S., J.R.G., H.A., H.L.R., L.V., and C.A.M. edited and revised manuscript; J.C.V., T.F.C., L.D.S., J.R.G., H.A., H.L.R., L.V., and C.A.M. approved final version of manuscript; T.F.C., L.D.S., H.L.R., and C.A.M. performed experiments.

REFERENCES

- Abdi H, Valentin D.** Multiple factor analysis. In: *Encyclopedia of Measurement Statistics*, edited by Salkind NJ. Thousand Oaks, CA: Sage, 2007, p. 657–663.
- Abdi H, Williams L.** Principal component analysis. *Wiley Interdiscip Rev Comput Stat* 2: 433–459, 2010.
- Calinski RB, Harabasz J.** A dendrite method for cluster analysis. *Commun Stat* 3: 1–27, 1974.
- Campbell TF, Dollaghan CA, Rockette HE, Paradise JL, Feldman HM, Shriberg LD, Sabo DL, Kurs-Lasky M.** Risk factors for speech delay of unknown origin in 3-year-old children. *Child Dev* 74: 346–357, 2003.
- Chermack GD, Schneiderman CR.** Speech timing variability of children and adults. *J Phon* 13: 477–480, 1986.
- Conti-Ramsden G, Crutchley A, Botting N.** The extent to which psychometric tests differentiate subgroups of children with SLI. *J Speech Lang Hear Res* 40: 765–777, 1997.
- Davis BL.** Clinical diagnosis of developmental speech disorders. In: *Phonological Disorders in Children*, edited by Kamhi AG, Pollock KE. Baltimore, MD: Paul Brookes, 2005, p. 3–21.
- Ding C, He X.** K-means clustering via principal component analysis. In: *Twenty-First International Conference on Machine Learning*. New York: ACM, 2004.
- Eguchi S, Hirsh IJ.** Development of speech sounds in children. *Acta Otolaryngol Suppl (Stockh)* 257: 1–51, 1969.
- Escofier B, Pagès J.** Multiple factor analysis. *Comput Stat Data Anal* 18: 121–140, 1990.
- Fielding A.** *Cluster and Classification Techniques for the Biosciences*. Cambridge, UK: Cambridge University Press, 2007.
- Fitch WT, Giedd J.** Morphology and development of the human vocal tract: a study using magnetic resonance imaging. *J Acoust Soc Am* 106: 1511–1522, 1999.
- Fukui T, Tsuruta M, Murata K, Wakimoto Y, Tokiwa H, Kuwahara Y.** Correlation between facial morphology, mouth opening ability, and condylar movement during opening-closing jaw movements in female adults with normal occlusion. *Eur J Orthod* 24: 327–336, 2002.
- Gan G, Ma C, Wu J.** *Data Clustering Theory, Algorithms, and Applications*. Alexandria, VA: American Statistical Association, 2007.
- Gildersleeve-Neumann CE, Kester ES, Davis BL, Pena ED.** English speech sound development in preschool-aged children from bilingual English-Spanish environments. *Lang Speech Hear Serv Sch* 39: 314–328, 2008.
- Goffman L.** Prosodic influences on speech production in children with specific language impairment and speech deficits: kinematic, acoustic, and transcription evidence. *J Speech Lang Hear Res* 42: 1499–1517, 1999.
- Goffman L, Gerken L, Lucchesi J.** Relations between segmental and motor variability in prosodically complex nonword sentences. *J Speech Lang Hear Res* 50: 444–458, 2007.
- Goffman L, Smith A.** Development and phonetic differentiation of speech movement patterns. *J Exp Psychol Hum Percept Perform* 25: 649–660, 1999.
- Goffman L, Smith A, Heisler L, Ho M.** The breadth of coarticulatory units in children and adults. *J Speech Lang Hear Res* 51: 1424–1437, 2008.
- Green JR, Moore CA, Higashikawa M, Steeve RW.** The physiologic development of speech motor control: lip and jaw coordination. *J Speech Lang Hear Res* 43: 239–255, 2000.
- Green JR, Moore CA, Reilly KJ.** The sequential development of jaw and lip control for speech. *J Speech Lang Hear Res* 45: 66–79, 2002.
- Green JR, Wilson EM, Wang YT, Moore CA.** Estimating mandibular motion based on chin surface targets during speech. *J Speech Lang Hear Res* 50: 928–939, 2007.
- Grillner S.** Control of locomotion in bipeds, tetrapods, and fish. In: *Handbook of Physiology. The Nervous System. Motor Control*. Bethesda, MD: Am. Physiol. Soc., 1981, sect. 1, vol. 2, part 2, p. 1179–1236.
- Hebb DO.** *The Organization of Behavior*. New York: Wiley, 1949.
- Holm A, Crosbie S, Dodd B.** Differentiating normal variability from inconsistency in children's speech: normative data. *Int J Lang Commun Disord* 42: 467–486, 2007.
- Jolliffe IT, Jones B, Morgan BJ.** Comparison of cluster analyses of the English personal social service authorities. *J R Stat Soc Ser A* 149: 253–270, 1986.
- Kent RD.** Anatomical and neuromuscular maturation of the speech mechanism: evidence from acoustic studies. *J Speech Hear Res* 19: 421–447, 1976.
- Kent RD, Forner LL.** Speech segment durations in sentence recitations by children and adults. *J Phon* 8: 157–168, 1980.
- Kleinow J, Smith A.** Potential interactions among linguistic, autonomic, and motor factors in speech. *Dev Psychobiol* 48: 275–287, 2006.
- Landtwing K.** Evaluation of the normal range of vertical mandibular opening in children and adolescents with special reference to age and stature. *J Maxillofac Surg* 6: 157–162, 1978.
- MacNeilage PF.** Motor control of serial ordering of speech. *Psychol Rev* 77: 182–196, 1970.
- Maner KJ, Smith A, Grayson L.** Influences of utterance length and complexity on speech motor performance in children and adults. *J Speech Lang Hear Res* 43: 560–573, 2000.
- Mirkin B.** *Clustering for Data Mining: A Data Recovery Approach*. London: Chapman and Hall, 2005.
- Moore CA, Ruark JL.** Does speech emerge from earlier appearing oral motor behaviors? *J Speech Hear Res* 39: 1034–1047, 1996.
- Pagès J, Husson F.** Multiple factor analysis with confidence ellipses: a methodology to study the relationships between sensory and instrumental data. *J Chemom* 19: 138–144, 2005.
- Riely RR, Smith A.** Speech movements do not scale by orofacial structure size. *J Appl Physiol* 94: 2119–2126, 2003.
- Robbins J, Klee T.** Clinical assessment of oropharyngeal motor development in young children. *J Speech Hear Disord* 52: 271–277, 1987.
- Shriberg LD, Allen CT, McSweeney JL, Wilson DL.** *PEPPER: Programs to Examine Phonetic and Phonologic Evaluation Records*. Madison, WI: Waisman Center, University of Wisconsin, 2001.
- Shriberg LD, Austin D, Lewis BA, McSweeney JL, Wilson DL.** The percentage of consonants correct (PCC) metric: extensions and reliability data. *J Speech Lang Hear Res* 40: 708–722, 1997.
- Shriberg LD, Campbell TF, Karlsson HB, Brown RL, McSweeney JL, Nadler CJ.** A diagnostic marker for childhood apraxia of speech: the lexical stress ratio. *Clin Linguist Phon* 17: 549–574, 2003.
- Shriberg LD, Fourakis M, Hall SD, Karlsson HB, Lohmeier HL, McSweeney JL, Potter NL, Scheer-Cohen AR, Strand EA, Tilkens CM, Wilson DL.** Perceptual and acoustic reliability estimates for the Speech Disorders Classification System (SDCS). *Clin Linguist Phon* 24: 825–846, 2010.
- Shriberg LD, Kwiatkowski J.** Developmental phonological disorders. I: A clinical profile. *J Speech Hear Res* 37: 1100–1126, 1994.
- Shriberg LD, Kwiatkowski J, Hoffmann K.** A procedure for phonetic transcription by consensus. *J Speech Hear Res* 27: 456–465, 1984.
- Shriberg LD, Lohmeier HL, Campbell TF, Dollaghan CA, Green JR, Moore CA.** A nonword repetition task for speakers with misarticulations: the Syllable Repetition Task (SRT). *J Speech Lang Hear Res* 52: 1189–1212, 2009.
- Shriberg LD, Lohmeier HL, Dollaghan CA, Campbell TF.** A nonword repetition task for speakers with speech sound disorders: the Syllable Repetition Task (SRT). In: *11th Meeting of the International Clinical Phonetics and Linguistics Association*, Dubrovnik, Croatia, 2006.
- Smith A, Goffman L.** Stability and patterning of speech movement sequences in children and adults. *J Speech Lang Hear Res* 41: 18–30, 1998.
- Smith A, Goffman L, Stark RE.** Speech motor development. *Semin Speech Lang* 16: ; quiz 98–89, 1995:87–98.
- Smith A, Zelaznik HN.** Development of functional synergies for speech motor coordination in childhood and adolescence. *Dev Psychobiol* 45: 22–33, 2004.
- Smith BL.** Temporal aspects of English speech production: a developmental perspective. *J Phon* 6: 37–67, 1978.
- Smith BL, Gartenberg TE.** Initial observations concerning developmental characteristics of labio-mandibular kinematics. *J Acoust Soc Am* 75: 1599–1605, 1984.
- Steeve RW, Moore CA.** Mandibular motor control during the early development of speech and nonspeech behaviors. *J Speech Lang Hear Res* 52: 1530–1554, 2009.
- Tan PN, Steinbach M, Kumar V.** *Introduction to Data Mining*. New York: Addison-Wesley, 2006.
- Thelen E.** Motor aspects of emergent speech: a dynamic approach. In: *Biological and Behavioral Determinants of Language Development*, edited by Krasnegor NA, Rumbaugh DM, Schiefelbusch RL, Studdert-Kennedy M. Hillsdale, NJ: Erlbaum, 1991, p. 339–362.
- Thelen E, Kelso J, Fogel A.** Self-organizing systems and infant motor development. *Dev Rev* 7: 39–65, 1987.
- Tingley BM, Allen GD.** Development of speech timing control in children. *Child Dev* 46: 186–194, 1975.

- van Lieshout PH, Moussa W. The assessment of speech motor behaviors using electromagnetic articulography. *The Phonetician* 81: 9–22, 2000.
- Venkatesh L. Speech movement characteristics of repetitive syllable production in children with speech disorders of unknown origin. In: *Speech and Hearing Sciences*. Seattle, WA: University of Washington, 2007, p. 130.
- Vorperian HK, Kent RD. Vowel acoustic space development in children: a synthesis of acoustic and anatomic data. *J Speech Lang Hear Res* 50: 1510–1545, 2007.
- Vorperian HK, Kent RD, Lindstrom MJ, Kalina CM, Gentry LR, Yandell BS. Development of vocal tract length during early childhood: a magnetic resonance imaging study. *J Acoust Soc Am* 117: 338–350, 2005.
- Walsh B, Smith A. Articulatory movements in adolescents: evidence for protracted development of speech motor control processes. *J Speech Lang Hear Res* 45: 1119–1133, 2002.
- Walsh B, Smith A, Weber-Fox C. Short-term plasticity in children's speech motor systems. *Dev Psychobiol* 48: 660–674, 2006.
- Wang K. CVAP: Cluster Validity Analysis Platform (Online). The MathWorks. <http://www.mathworks.com/matlabcentral/fileexchange/14620-cvap-cluster-validity-analysis-platform-cluster-analysis-and-validation-tool>.
- Westbury JR. Mandible and hyoid bone movements during speech. *J Speech Hear Res* 31: 405–416, 1988.
- Wiig EH, Secord W, Semel E. *Clinical Evaluation of Language Fundamentals—Preschool*. San Antonio, TX: Psychological Corporation, 1992.
- Xie H, Andreae P, Zhang M, Warren P. Detecting stress in spoken English using decision trees and support vector machines. In: *Proceedings of the Second Workshop on Australasian Information Security, Data Mining and Web Intelligence, and Software Internationalisation*. Dunedin, New Zealand: Austral. Comput. Soc., 2004, vol. 32.

