Published in final edited form as:
Top Lang Disord. 2011 ; 31(2): 168–184. doi:10.1097/TLD.0b013e318217b875.

Phonetic variability in residual speech sound disorders: Exploration of subtypes

Jonathan L. Preston, PhD
Haskins Laboratories & Department of Communication Disorders, Southern Connecticut State University

Laura L. Koenig, PhD
Haskins Laboratories & Department of Communication Sciences and Disorders, Long Island University – Brooklyn Campus

Abstract

Purpose—To explore whether subgroups of children with residual speech sound disorders (R-SSDs) can be identified through multiple measures of token-to-token phonetic variability.

Method—Children with R-SSDs were recorded during a rapid multisyllabic picture naming task and an oral diadochokinetic task. Transcription-based and acoustic measures of token-to-token variability were derived. Articulation accuracy and general indices of language skills were also measured.

Results—Low correlations were observed between transcription-based and acoustic measures of phonetic variability, and among the acoustic measures themselves. Children who were the most variable on one measure were not necessarily highly variable on other measures. Transcription-based measures of variability were associated with language skills.

Conclusions—Measures of phonetic variability did not identify children in the sample as consistently high or low. Data do not support the notion that clear subgroups based on phonetic variability can be reliably identified in children with R-SSDs. The link between highly variable phonetic output (quantified by transcription-based measures) and lower language skills requires further exploration.

Keywords

acoustic analysis; phonetic variability; residual speech sound disorders

INTRODUCTION

Individuals with speech sound disorders (SSDs) have clinically significant errors in production of the sounds of their native language. Those with residual speech sound disorders (R-SSDs) have speech sound errors of unknown origin that persist beyond the age of typical acquisition (Shriberg, 1994, 2009), so are generally age 9 and above. Nearly all English-speaking children with R-SSDs are found to misarticulate liquids, sibilants, or both; among the liquids, errors are often observed on /ɹ, ɝ, ɚ/ and sometimes also /l/, and sibilant errors usually include /s, z/ and sometimes /ʃ, ʒ, tʃ, dʒ/ (e.g., Irwin, Knight, & Oltman, 1974; Shriberg, 2009; Shriberg, Gruber, & Kwiatkowski, 1994). Liquids and sibilants are generally acquired later by English-speaking children (e.g., Smit, Hand, Freilinger, Bernthal,

---

Please send correspondence to: Jonathan Preston, Haskins Laboratories, 300 George St, Ste 900, New Haven, CT 06511, preston@haskins.yale.edu. Fax: 203-865-8963 Phone: 203-865-6163.

& Bird, 1990), and may be more difficult to master because of the precise motor control required to produce the articulatory gestures of these sounds.

Although several studies have explored subtypes among preschool and young school-age children with SSD, less is known about subtypes of R-SSDs. This may be because R-SSDs, not surprisingly, are less prevalent than preschool SSDs. Whereas the prevalence of SSD at age 3 is estimated to be as high as 15.6% (Campbell et al., 2003), that of R-SSD may be approximately 1–5% (Culton, 1986; Irwin, Huskey, Knight, & Oltman, 1974; Irwin, Knight, et al., 1974; Shriberg, 1994). Shriberg and colleagues (Shriberg, 1994, 2009; Shriberg, Austin, Lewis, McSweeny, & Wilson, 1997a) have described classification of R-SSDs by the primary sound errors (e.g., /ɹ/ and /s/), by etiology (e.g., genetic vs. environmental), and/or by speech history (i.e., whether the child had a history of SSD at a younger age that has not yet resolved, or whether the R-SSD is evident following otherwise typical speech development). Although it can be difficult to reliably subtype by etiology or speech history (Dodd, 2005), few other subtyping schemes have been proposed for R-SSDs.

Phonetic Variability and Speech Motor Control

Many studies of typical speech development have documented an extended trend, lasting into the adolescent years, of decreasing variability in repeated productions of a word or sound (e.g., Eguchi & Hirsh, 1969; Kent & Forner, 1980; Lee, Potamianos, & Narayanan, 1999; Munson, 2004; Ohde, 1985; Sharkey & Folkins, 1985; Smith & McLean-Muse, 1986; Walsh & Smith, 2002). Although some variability is thus expected in development, high token-to-token variability at the word and phoneme levels has been described as a clinical marker of SSD or of subtypes of SSD (Bradford & Dodd, 1996; Dodd, 2005; Holm, Crosbie, & Dodd, 2007).

Inconsistency in repeated productions of words has specifically been noted as a feature of inconsistent disorder (a subtype described by Dodd [2005] that involves high token-to-token inconsistency but normal nonverbal oral motor skills) or of childhood apraxia of speech (CAS), a subtype of SSD that has received much attention but can be difficult to diagnose reliably (see ASHA, 2007). Children with inconsistent disorder or CAS are believed to have deficits in motor planning or programming (ASHA, 2007; Nijland, Maassen & van der Meulen, 2003), which would predict inconsistent phonetic output. Some studies suggest that young children with inconsistent speech errors or CAS tend to progress slowly in therapy (e.g., Forrest Dinnsen, & Elbert, 1997), and therefore might be likely to have unresolved speech errors that persist past age 9 (i.e., R-SSD). In most reports, children with inconsistent productions or CAS show evidence of other co-occurring deficits such as lower language skills or greater severity of speech impairment (Lewis, Freebairn, Hansen, Iyengar, & Taylor, 2004; Tyler, Williams & Lewis 2006), which may indicate system-wide deficits in speech and language.

It is unclear, however, whether the patterns associated with inconsistent output in preschool or young school-age children are also evident in R-SSDs, where speech systems are more developed. Inconsistent productions have been observed in some case studies of individuals with R-SSD (Hall, 1989; Shuster, Ruscello, & Haines, 1992; van Lieshout, Merrick, & Goldstein, 2008), but exploration of phonetic variability in a larger cohort is needed. Because older children generally show less variable speech production than younger children, phonetic variability may have limited utility for detecting unique differences among children with R-SSD. In a recent epidemiological report of children referred for speech impairments, Broomfield and Dodd (2005) found that none of the seven children who were 11 years or older met criteria for “inconsistent disorder”. Therefore, inconsistency may not be a common trait among older children. Thus, the present study explores whether token-to-token variability can be used to identify subgroups of children with R-SSDs, and whether phonetic variability at later ages relates to other clinically-relevant characteristics, including language skills and the sound classes that are misarticulated. Specifically, we use both transcription-based and acoustic measures to determine if highly variable speech is evident across the speech systems of some children with R-SSD but not others.

Quantifying Phonetic Output—Transcription-based measures of token-to-token variability have been used for subtyping SSDs or to predict growth in speech accuracy in preschoolers with SSD (Dodd, 2005; Marquardt, Jacks, & Davis, 2004; Thoonen, Maassen, Gabreels, Schreuder, & de Swart, 1997; Tyler & Lewis, 2005; Tyler, et al., 2006). However, transcriptional measures can be characterized by biases and questionable reliability, and may lack sensitivity to phonetic variability in older children who have more mature speech systems. Data presented by Dodd, Holm, Crosbie and McCormack (2005) showed that preschoolers with SSD who were classified as inconsistent from token-to-token based on transcriptional measures were less consistent in vowel formant values than preschoolers with SSD whose errors were consistent, providing some evidence that transcriptional and acoustic measures of variability are related.

Another method of evaluating the development of speech motor control is the study of speech rate in various tasks. For example, diadochokinetic (DDK) tasks, which require rapid production of syllables or syllable sequences, are commonly used to evaluate speech motor functioning in children (Fletcher, 1972; Thoonen, Maassen, Wit, Gabreels, & Schreuder, 1996). McNutt (1977) subgrouped children with R-SSD by the phoneme in error (/ɹ/ or /s/), and compared these subgroups to typically speaking children on a DDK task. He found that both subgroups of R-SSD were slower in repetition of syllables than typically speaking controls, which was interpreted to reflect differences in speech motor functioning. In contrast, Preston and Edwards (2009) found that children with rhotic errors were not significantly slower than typically speaking peers on rapid productions of /pʌtʌkʌ/. However, the children with R-SSDs did appear to be more variable, on average, than controls, as evidenced by a greater variety of error types in the syllable sequence (i.e., they were more likely to produce several different types of errors), as well as more errors overall in place, manner, and voicing. Flipsen (2003) also evaluated speaking rate in children with R-SSDs, considering conversational speech and target words embedded in phrases. Younger (9-year-old) and older (12–16-year-old) children with R-SSDs were compared to age-matched control children whose speech errors had resolved. Whereas the speech rate of 9 year olds with R-SSDs did not differ from the controls on either task, the older children with R-SSDs were significantly slower than controls in producing embedded words (but not conversational speech). Such rate differences may implicate poor speech motor control and/or language formulation deficits among older children with R-SSDs (Flipsen, 2003). Finally, picture naming has also been shown to be slower and less accurate among children with R-SSD (Preston & Edwards, 2009). Thus, an assessment of variability in duration and accuracy in both single words and DDK sequences may also aid in the understanding of individual differences in children with R-SSDs.

Language skills

There is evidence that language impairments often co-occur with preschool SSD; however, little is known about specific relationships between phonetic patterns and language skills in R-SSD. Both Flipsen (2003) and Preston and Edwards (2009) found that children with R-SSDs, as a group, made more errors in lexical choice, morphology, and syntax than controls in a sentence repetition task. Additionally, Gross, St. Louis, Ruscello, and Hull (1985) observed poorer expressive language (grammatical) skills among first, third, fifth and seventh graders with multiple errors (errors on final consonants, not including their errors on ɹ, s, l/) relative to age-matched children with errors limited to ɹ, s, l/.

Yet, questions remain about the relationships between honetic variability and language skills in R-SSDs. It might be the case that poorly- or incompletely-formed linguistic representations yield inconsistent phonetic output as language is mapped to speech, or that inconsistent speech production interferes with lexical and morpho-syntactic development (cf. Shriner, Holloway & Daniloff, 1969). Seeff-Gabriel, Chiat and Dodd (2005) compared 4–6 year olds with consistent speech errors and those with inconsistent speech errors on repetition of long and complex sentences; they found that children with inconsistent errors were more likely to make errors in repetition of content words, function words and inflections than children with consistent speech sound errors. Similarly, Lewis et al. (2004) observed lower scores on clinical measures of imitation and receptive and expressive language in preschoolers with symptoms of CAS (who were presumably more phonetically variable) compared to children with SSD without such symptoms. Differences in language skills were also evident when the children were followed-up at ages 8–10. Thus, another goal of this work is to explore whether inconsistent phonetic output is related to deficits in sentence imitation and vocabulary in children with RSSDs.

The present study takes an exploratory approach to subtyping R-SSDs. That is, we make no *a priori* assumptions about the subgroup classification of participants; instead, we investigate whether clear subgroups of highly consistent or inconsistent speakers are evident in the data obtained from both transcription-based and acoustic measures. If subtle speech motor control difficulties are pervasive throughout a child's speech system, multiple measures of variability (both acoustic and transcription-based) should be strongly correlated, and individuals who are relatively high in one index of phonetic variability should be relatively high in another. Since prior work has suggested that sound classes in error (e.g., sibilants and liquids) might differentiate subgroups of R-SSD (e.g., McNutt, 1977), we examine whether errors on one or both of these sound classes relate to consistency of speech production. Additionally, we evaluate how phonetic variability relates to performance on clinical measures of language functioning.

METHODS

Participants and general procedures

Twenty native English-speaking children ages 9;2 –15;5 years (mean 12;1, 12 males, 8 females) from central New York and southern Connecticut were referred by clinicians or parents. All demonstrated at least one sound with perceptually identifiable errors in connected speech and in single words (minimally, ɹ/ or /s/, and often other liquids and sibilants). Therapy histories varied, ranging from no intervention to continued intervention since age 2. Although some parents reported that CAS had been suspected, this was not used as a grouping criterion due to the lack of agreed upon criteria (see ASHA 2007 for further discussion of this matter). Participants had nonverbal IQ scores above 75 as reported by the referring clinician or based on administration of the nonverbal composite of the Differential Ability Scales (Elliott, 1990). Evaluation sessions with each child and the first author were recorded in a quiet room with either an Olympus WS-331M digital voice recorder recorded on high quality mode (with no low-cut filter, sampled at 44kHz), or a Shure WH20 head-mounted microphone fed into a Rolls MX 54s Pro Mixer Plus and recorded on a Dell Inspiron 8600 laptop (sampled at 22kHz in Praat), providing adequate temporal resolution for the measures of interest (see below). These recorded data were used for subsequent transcriptions as well as acoustic analyses.

Each participant's accuracy on sibilants /s, z, ʃ, ʒ, tʃ, dʒ/ and liquids /l, ɹ/ was computed based on phonetic transcriptions of his/her productions during a 64-item picture naming task (see Appendix A in Preston & Edwards, 2007), which provided opportunity for producing 68 sibilants and 50 liquids. The sample included many multisyllabic words and consonant clusters, including every English consonant (except /h/) at least twice . Each participant was below 70% accurate on at least one of these broad classes of sounds on this single word task. These children were beyond the typical age of acquisition of these sounds (Smit et al., 1990), and were therefore considered to have R-SSDs. Percent Consonants Correct (Shriberg & Kwiatkowski, 1982; Shriberg et al., 1997b) was used to quantify phonetic accuracy based on the transcriptions of these 64 words.

Language assessment included two common clinical measures. The Peabody Picture Vocabulary Test—III (Dunn & Dunn, 1997) evaluates single word receptive vocabulary. The Recalling Sentences subtest of the Clinical Evaluation of Language Fundamentals-4 (Semel, Wiig, & Secord, 2003) evaluates verbal and grammatical working memory in imitated sentences that increase in length and syntactic complexity.

Multiple phonetic measures were derived from rapid naming and diadochokinesis tasks (see Table 1). Measures were selected that could be reliably obtained and that have been reported elsewhere in studies of development, variability, SSDs, and/or R-SSDs. The transcriptional measures were the Error Consistency Index (ECI) (Tyler & Lewis, 2005; Tyler, Lewis, & Welch, 2003; Tyler et al., 2006); Total Token Variability (TTV; Marquardt et al., 2004); and the number of different forms produced in 40 productions of /pʌtʌkʌ/ during a diadochokinetic task (Preston & Edwards, 2009). Acoustic measures were carried out in Praat (Boersma & Weenink, 2008) and included word and trisyllable durations, vowel formants, and voice onset time (VOT, Lisker & Abramson, 1964). Duration measures reflect speech rate and have been widely used in studies of speech motor control, development, and SSD (e.g, Flipsen, 2003; Kent & Forner, 1980; Lee et al., 1999). Formants provide phonetic information for vowels, and have been evaluated in past studies of CAS (Nijland et al., 2003) and inconsistent phonological disorders (Seeff-Gabriel et al., 2005); moreover, considerable normative data are available (Eguchi & Hirsh, 1969; Hillenbrand et al., 1995; Kent & Forner, 1979; Lee et al., 1999). Finally, VOT, the interval between plosive release and onset of voicing, provides a measure of stop voicing and aspiration. VOT and its variability have been studied extensively in development (e.g., Kent & Forner, 1980; Koenig, 2001; Ohde, 1985) and in adults with motor speech disorders (see Auzou et al., 2000; Kent & Kim, 2003).

Phonetic variability in a 64-item picture naming task

As a global measure of consonant variability, the ECI was obtained from participants' productions of the 64-item picture naming task described above. The ECI is the sum of different error forms produced for all consonants combined. For example, if target /s/ was realized as [t], [d], and [ʃ], and /f/ was realized as [b] and [p], with no other phonemic errors produced in the sample, the ECI would be 5. Omissions were counted as one type of error form. Additionally, a broad category of "distortion" was allowed to quantify one variant of each target consonant, but different allophonic variations (e.g., labialization and derhoticization of ɹ) were not considered different substitutes because of the inherent difficulty with obtaining reliable narrow transcriptions in disordered speech (Shriberg & Lof, 1991). Although these children's errors were primarily on sibilants and liquids, occasional errors on other sounds were observed.

Phonetic variability in rapid multisyllabic picture naming

Word-level token-to-token consistency for multisyllabic words was assessed in a rapid naming (RN) task described by Preston and Edwards (2009) in which participants were instructed to quickly name the pictures on a page. The target words elephant, umbrella, strawberries, helicopter, thermometer, and *spaghetti* were named once initially, then elicited 4–5 times each in the RN task. This task was chosen because (a) it involved repetitions that could be analyzed for token-to-token consistency, (b) it included rapid production of multisyllabic words which might tax the speech motor system, and (c) it has shown large differences between children with R-SSD and typically speaking peers with respect to the duration of naming and the number of words with speech sound errors (Preston & Edwards, 2009). In the present work, we ask whether phonetic variability in repeated productions of words in an RN task might inform our understanding of within-group differences among children with R-SSD.

Transcriptional measures of variability on the RN task—Total Token Variability (TTV) was computed for each of the six words repeated in the RN task. The TTV represents the proportion of spoken renditions of a word that are unique (i.e., unlike any other rendition of the word spoken by that child), and is calculated as follows: TTV = [# variants –1] / [# tokens – 1]. To obtain a single measure for each child, the average TTV for all of the six words was used. Phoneme substitutions, additions, or omissions were counted as unique variants. Following Marquardt et al. (2004), minor deviations in vowel accuracy or allophonic variations (e.g., aspirated vs. unaspirated stops) were not considered unique variants, nor were patterns that would readily be accounted for by rapid speech (e.g., lenition processes such as /b/ → [β]). Allophonic differences across tokens were considered unique variants only if the allophone involved alternation between a perceived correct production and a clinically significant distortion of a sound (e.g., alternation between correct /s/ and lateralized /s/).

Acoustic measures of variability on the RN task—From the RN task, a researcher who was blind to the results of the transcription-based measures extracted word durations for all renditions of the 6 words, along with formant values (F1 and F2) of 8 vowels: /ɛ/ in *elephant, umbrella, helicopter*, and *spaghetti*; /a/ in *helicopter* and *thermometer*, /ɔ/ in *strawberries*, and /i/ in *spaghetti* (see Tables 2A and 2B for details). Criteria for word durations and formant measurement locations were designed to ensure reliability in the face of coarticulatory and allophonic variation. For example, word offset in *elephant* did not depend on whether the final stop was realized as a released /t/, an unreleased /t/, or a glottal stop. For *helicopter* and *thermometer*, word onset was defined according to voicing because frication noise for /h/ and /θ/ is often weak or absent altogether. Formant measures of /ɛ/ in *helicopter* and *elephant* were taken early in the vowel to limit the effects of the following liquid. In the case of *umbrella*, the maximum F2 was chosen to yield the most "/ɛ/-like" value possible at a reliable location between the two liquids. Exclusion criteria for the formant measures were overt lexical substitutions (e.g., *pasta* produced for *spaghetti*) and cases where the signal quality did not permit formant values to be determined. Data loss because of extraneous noise or inability to resolve the formants was minimal: 3.4% for formant values and 4.8% for word durations.

For each of the 8 target vowels, the mean F1 and mean F2 were obtained for each child from his/her repeated productions. Then, an estimate of token-to-token variability was computed by calculating the Euclidean distance (in two dimensions) from the mean F1 and F2 for all productions of the vowel. Essentially, the F1–F2 means represented x, y coordinates in two-dimensional space, and Euclidean distances represented straight lines between the mean and each token of the vowel. Thus, high token-to-token variability (less acoustic consistency) in vowel production was represented by large Euclidean distances. The average Euclidean distance over all 8 vowels was used to estimate a speaker's overall acoustic variability in vowel production.

To estimate variability in word duration, the coefficient of variation (COV) was computed for each word and participant. The COV is the ratio of the standard deviation to the mean and is therefore a unitless estimate (percentage) of variability. The weighted average COV was computed for each subject from the six target words as the measure of variability in word duration (with COVs weighted based on the number of productions of each word).

Phonetic variability in an oral diadochokinetic task

The oral diadochokinetic (DDK) task required each participant to repeat the syllable sequence /pʌtʌk/ʌ as quickly as possible. At least 10 sequences of the trisyllable were elicited per trial, with up to 6 trials per participant.

Transcriptional measures of variability on the DDK task—For 40 trisyllable attempts (four trials of 10 consecutive productions), phonetic transcriptions were used to capture deviations in place, manner, and voicing of the three target consonants, and the total number of different error forms of /pʌtʌkʌ/ over the 40 productions was derived (see Preston & Edwards, 2009). Thus, if a child produced [pʌkʌkʌ], [bʌkʌkʌ] [pʌtʌpʌ] and [pʌkʌ], the DDK error variability score was 4. To maximize reliability, two transcribers listened to the productions to achieve consensus (cf. Shriberg, Kwiatkowski & Hoffman, 1984).

Acoustic measures of variability on the DDK task—Two acoustic measures were taken from the DDK task: trisyllable duration and VOT for /p/ and /k/. Trisyllable duration was measured from the burst of the first stop to the offset of voicing in the third syllable or the consonantal closure for the subsequent trisyllable (whichever came first). Measures were only made in cases where the production was perceived to have 3 syllables (no added or deleted syllables; errors in consonant voicing, place or manner were allowed). For VOT, exclusion criteria were cases where the release burst could not be identified (e.g., spirantization of the stop, background noise), or where the place of articulation for the first or third consonant did not match the target. VOT was not attempted for /t/ because it is frequently realized as [ɾ]. In running speech, intervocalic stops may often show voicing continuing into the stop closure from the preceding vowel. Such cases of closure voicing were ignored here, so that all VOT values were positive. On average, 79.6% and 62.7% of productions were measurable for /p/ and /k/ respectively, corresponding to a total of 673 tokens of /p/ and 530 tokens of /k/. Phonetic variability was quantified for each participant by computing the COV for trisyllable durations and for /p/ and /k/ VOTs.

Reliability

A transcriber blind to purposes of the analysis transcribed the responses from the 64 item naming task for 9 randomly selected participants to obtain ECI scores. Average point-by-point consonant agreement (including rhotic vowels) was 87.3% (range 81.5 –92.3), which is similar to other studies (Shriberg et al., 1997b). ECI scores from the reliability transcriptions correlated highly with the ECI scores derived by the first author ($r = 0.98$, $p < .001$; mean difference 2.7, SD 3.1). For TTV, the second author independently phonetically transcribed six participants on the RN task. TTV agreement did not reach an acceptable level of reliability with the transcriptions done by the first author ($r = 0.55$, $p = 0.001$, mean difference 0.07, SD 0.20), presumably because of the inherent difficulty in transcribing rapid, disordered speech, and because disagreement on any single phoneme would impact the overall TTV score (n.b., prior studies have not reported reliability on the TTV). To address this issue, both authors independently transcribed all 18 participants and the average TTV from the two transcribers was used.

For each acoustic measure, a minimum of 100 tokens were re-analyzed for each measure via both inter- and intra-rater reliability. The results (Table 3) show high reliability for all acoustic measures, with small mean differences between the original and re-measured values. Intra-rater reliability was computed by the second author, with the new measures obtained at least a week after the original ones. Inter-rater reliability was also obtained for all measures, with the first author measuring DDK values and a graduate student in speech-language pathology measuring the RN data In the case of RN data, the student was given the measurement protocol as part of a training exercise for future acoustic analysis work; the initial set of measurements performed on data from 6 children, even prior to addressing frank measurement errors, showed strong reliability. DDK VOTs and durations were also highly reliable.

RESULTS

Table 4 provides summary statistics of the variables explored in this study. From the PCC scores, it is clear that these children have R-SSDs, and that there is a range of phonetic accuracy. ECI scores are substantially lower than the values reported by Tyler et al. (2003) with preschoolers, indicating that, as would be expected, adolescents are not as variable (or as inaccurate) as preschoolers. PPVT and Recalling Sentences scores extend from low performance to above average, indicating that there is substantial range in the data to examine high and low skill. Because age did not correlate with any of the measures of phonetic variability, no statistical adjustment for age was made.

Graphical display of data is used to evaluate whether subgroups of highly consistent and/or inconsistent children existed among the children with R-SSDs. This is because of the limited sample size, and because outliers or clustering of participants into subgroups can be observed with graphical inspection of data, but might be missed when data are collapsed for statistical analyses. If clear subgroups are present, we would expect extreme outliers or bimodal distributions in the data. Inspection of the histograms in Figure 1 does not readily support this. Indeed, the measures of variability in Figure 1 appear to be quite continuous in nature with little evidence of clustering at the high or low ends of the phonetic variability spectrum. Although some outliers are evident (e.g., DDK duration, DDK /p/ VOT), the same child is not the outlier on these measures.

It is apparent that two of the transcription-based measures are highly correlated: word-level token-to-token consistency (TTV) and overall error consistency (ECI). Inconsistency, as measured by both ECI and TTV, is associated with vocabulary and sentence imitation. To further examine whether sounds in error are associated with language skills and phonetic variability, each child was classified as misarticulating sibilants, liquids or both based on a criterion of below 90% accuracy on these sounds on the 64 item picture naming task. This resulted in two participants with errors only on sibilants, seven whose errors were limited to liquids, and 11 with errors on both of these sound classes. The overall pattern is for children with errors on both liquids and sibilants to score lower on these language measures than children whose errors were restricted to one sound class. Additionally, the ECI-TTV plot shows that children who misarticulate one sound class tend to have lower ECI and TTV scores than children who misarticulate both sibilants and liquids. Mann-Whitney tests comparing children whose errors were restricted to one sound class (liquids or sibilants) to children whose errors involved both sound classes indicated revealed marginally significant group differences on TTV ($p = 0.035$ one-tailed) and stronger group differences for ECI ($p = 0.0015$ one-tailed).

In marked contrast, the acoustic measures of token-to-token consistency show very little relationship with each other, or with the transcription-based measures. This is the case even for measures derived from the same speaking task: TTV, vowel formant, and word duration variability measures were all obtained from the RN task, but failed to show significant relationships with each other. The plot of vowel formant variability and COV of word durations shows that participants who were relatively consistent in their vowel productions were not necessarily the most consistent in word durations on the RN task. Similarly, on the DDK task, variability in VOTs and trisyllable durations were not strongly related with one another or with the number of error forms perceived. Finally, unlike the transcriptional measures, the acoustic data do not provide converging evidence of a relationship between token-to-token inconsistency and language skills. In fact, variability in vowel formants shows an unexpected trend, with greater variability in formants associated with better vocabulary.

Figure 2 presents each participant's ranks on the measures of phonetic variability. Higher ranks represent greater token-to-token variability. Though some speakers have multiple values that are relatively low (e.g., participants 1, 5, and 9) or relatively high (e.g., participant 7's acoustic measures), it cannot be said that these speakers cluster at very low or high ends of the scale. For example, Participant 7, whose acoustic variability was rather high, is relatively low on scores of transcriptional variability. Thus, the measures as a set fail to converge on the same children, providing little evidence that individual children can be consistently identified as high or low in phonetic variability.

DISCUSSION

The present study was an exploratory investigation of phonetic variability in R-SSD using both transcription-based and acoustic measures. It was presumed that if some children were indeed highly variable from token to token, then both acoustic and transcription-based measures would provide converging evidence of such variability. Although the data showed some correspondences among the transcription-based measures, this was not universally true. Moreover, the acoustic measures did not correlate among themselves, nor with the transcription-based measures.

Moderate correlations among two of the transcription-based measures, the ECI and TTV, provide some support for the notion that variability in one domain (variable realizations of the English consonants, the ECI) can be associated with variability in another domain (token-token variability in repeated productions of words, TTV). This relationship is as expected: children with multiple substitutions for target phonemes would be more likely to have multiple error forms when repeatedly naming a word. Similarly, children who misarticulated multiple sound classes (sibilants and liquids) were observed to be more phonetically variable than children who misarticulated a single sound class based on the transcription measures (see ECI-TTV plot in Figure 1). This may be because highly variable phonetic output is a sign of delayed/deficient speech motor control; thus, inaccurate speech and highly variable output are expected to co-occur (cf. Tyler et al., 2006).

Similar to some past work, there was a trend for lower language skills to be associated with less accurate output (cf. Gross et al., 1985) and greater phonetic variability based on transcription-based measures (cf. the CAS group studied by Lewis et al., 2004 and the children with inconsistent phonological disorder described by Seeff-Gabriel et al., 2005). Highly variable output may hinder the mapping of linguistic information to a phonetic form, or vice-versa; that is, variable feedback (e.g., auditory/proprioceptive) during production of words or sentences might impede the formation of stable linguistic representations (cf. Shriner et al., 1969), and/or unstable linguistic representations might yield variable phonetic output. Further exploration is required to understand the nature of the association between language skills and phonetic variability among children with R-SSD.

Despite the few observed correlations, the data on the whole do not show widespread correspondences among variability measures taken across domains. Neither of the transcription-based measures of variability (TTV or ECI) was strongly related to the transcription-based variability score from the DDK task. Phonetic variability as determined by the acoustic measures did not correlate with articulation accuracy or language skills (except for the correlation in the unexpected direction between vowel formants and language skills). The lack of correlation between the acoustic and transcription-based measures of phonetic variability suggests that the segmental variability perceived by the transcribers was not associated with the acoustic variability in vowel formants, VOT, or word and trisyllable durations. Perhaps more surprising, the acoustic measures of phonetic variability (the most reliable measures) were not significantly related to one another. Even different durational measurements were not strongly related: Children who were the most variable in DDK trisyllable durations and VOTs were not necessarily the most variable in word durations.

Some studies of variability in typical development have also reported poor within-speaker correspondences among instrumental measures of speech production variability. For example, children who are highly variable in speaking sound pressure level are not necessarily variable in measures of subglottal pressure or respiratory system volumes (Stathopoulos, 1995), and some children who show aerodynamic variability in one fricative (e.g., /s/) may be quite consistent on others (Koenig, Lucero, & Perlman, 2008). The current data similarly indicate that children who were the most inconsistent in one phonetic feature (e.g., VOT) were not necessarily the most inconsistent in others (e.g., vowel formants).

With respect to the current consideration of speech production variability in R-SSDs, additional factors may contribute to the lack of convergence among the measures of phonetic variability. First, it is possible that the measures used were not sufficiently sensitive to phonetic variability. For example, the TTV values did not yield strong reliability and may therefore be weak indicators of phonetic variability. Although all of the acoustic measures chosen for this study have been widely-used in clinical populations and in developmental studies (including those that have traced token-to-token variability across ages), other acoustic measures, perhaps including dynamic measures (e.g., formant transitions) might reveal interrelationships (see, e.g., Sussman et al., 2000). Similarly, one might argue that our acoustic measures under-represent some forms of variability. For example, DDK VOTs were not measured for stops with errors in place of articulation, and DDK durations were only measured for instances in which three syllables were perceived.

Alternatively, the lack of convergence among measures may simply indicate that phonetic variability is not necessarily pervasive throughout the speech systems of a subgroup of children with R-SSDs. Perhaps the extensive variability observed among some preschoolers with SSD can no longer be reliably observed at older ages. Had there been unique subgroups of highly variable or highly consistent children with R-SSD, we might have observed outliers that clustered together at either end of the continuum of phonetic variability. Although some of these children (primarily participants with errors on both liquids and sibilants) had what was reported by parents or clinicians as suspected CAS, the data did not converge on identifying these participants as excessively variable from token to token across measures. (It should be acknowledged that token-to-token variability is not the only symptom of CAS.)

High phonetic variability may reflect breakdowns in consistently planning/ programming coordinated speech movements (Dodd, 2005; ASHA, 2007). However, other psycholinguistic processes, including the quality of phonological representations, access to those representations, or the precise execution of speech movements, may be the locus of some R-SSDs (e.g., Dodd, 2005; Pascoe, Stackhouse & Wells, 2006). The transcriptional measures may capture a phonological level of representation/processing whereas at least some of the acoustic measures (e.g., formant values; DDK trisyllable durations) may better reflect phonetic or motoric processes. Without undertaking a detailed discussion of the nature of the phonetics-phonology interface, we can note that measurements of variability at different processing levels need not inherently be dissociated. It would be reasonable to expect that system-wide deficits in "higher level" phonological representations or processes would be expressed in the acoustic phonetic signal. However, the reverse is not necessarily true (i.e., lower-level phonetic deficits may not necessarily be revealed in high-level phonological processes), which could exlain the lack of association among some of the measures.

A clear limitation of the current study is the sample size, which makes it difficult to draw clear conclusions about subgroups, or to entirely rule out the possibility that some relatively rare children may exist who are widely variable in all aspects of speech production. Indeed, few large-scale databases of R-SSD exist, making the study of this population a challenge, particularly when searching for subgroups. Because many factors may be related to phonetic variability in R-SSD (e.g., intervention, etiology), longitudinal data may help us to better understand whether excessive variability on any given task or measure is stable across time, or whether phonetic variability itself is variable.

The present study appears to provide the first exploration of phonetic variability among a cohort of children with R-SSDs. The results make it difficult to argue that unique subgroups of phonetically variable children with R-SSDs can be reliably identified. Whereas the bases of phonetic variability have previously been studied in neurogenic populations (e.g., Kent et al., 2000; Kent & Kim, 2003; Shuster & Wambaugh, 2008), the neurobiological bases of phonetic variability in childhood SSD remain understudied (cf. ASHA, 2007), and a better understanding of neurobiological and psycholinguistic factors may aid in our understanding of possible subtypes. If "phonetic variability" is a dimension on which some children with R-SSD do show clear differences, it will be critical to further specify (theoretically, empirically, and biologically) how such variability arises, under what conditions it is observed, and whether it would be important for differentiating these individuals clinically.

Acknowledgments

Support for preparation of this manuscript was provided by NICHD grant T32HD7548 to Haskins Laboratories (C. Fowler, PI). Thanks to Heather Ramsdell and to Mary Louise Edwards for assisting with transcriptions; Dyala Sophia Eid for assistance with acoustic data processing (supported by a Graduate Assistantship from Long Island University, Brooklyn Campus); and Carol Fowler for comments on an earlier version of this manuscript.

REFERENCES

ASHA. Childhood Apraxia of Speech [Technical Report]. 2007. Available from www.asha.org/policy

Auzou P, Ozsancak C, Morris RJ, Jan M, Eustache F, Hannequin D. Voice onset time in aphasia, apraxia of speech and dysarthria: a review. Clinical Linguistics & Phonetics. 2000; 14(2):131–150.

Boersma, P.; Weenink, D. *Praat v* 5.0.34. 2008. www.praat.org

Bradford A, Dodd B. Do all speech-disordered children have motor deficits? Clinical Linguistics and Phonetics. 1996; 10(2):77–101.

Broomfield, J.; Dodd, B. Epidemiology of speech disorders. In: Dodd, B., editor. Differential diagnosis and treatment of children with speech disorders. 2nd ed.. Whurr Publishers; London: 2005. p. 83-99.

Campbell TF, Dollaghan CA, Rockette HE, Paradise JL, Feldman HM, Shriberg LD, Sabo DL, Kurs-Kasky M. Risk factors for speech delay of unknown origin in 3-year-old children. Child Development. 2003; 74(2):346–357. [PubMed: 12705559]

Culton GL. Speech disorders among college freshmen: a 13-year survey. Journal of Speech and Hearing Disorders. 1986; 51:3–7. [PubMed: 3945057]

Dodd, B., editor. Differential diagnosis and treatment of children with speech disorder. 2 ed.. Whurr Publishers; Philadelphia: 2005.

Dodd, B.; Holm, Alison; Crosbie, S.; McCormack, P. Differential diagnosis of phonological disorders. In: Dodd, B., editor. Differential diagnosis and treatment of children with speech disorder. 2nd ed.. Whurr Publishers; Philadelphia: 2005. p. 44-70.

Dunn, LM.; Dunn, LM. Peabody Picture Vocabulary Test. 3rd ed.. AGS; Circle Pines, MN: 1997.

Eguchi S, Hirsh IJ. Development of speech sounds in children. Acta Otolaryngologica. 1969:57.

Elliott, CD. Differential Ability Scales. Psychological Corporation/Harcourt Brace; San Antonio, TX: 1990.

Fletcher SG. Time-by-count measurement of diadochokinetic syllable rate. Journal of Speech and Hearing Research. 1972; 14(4):763–770. [PubMed: 4652397]

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

Forrest K, Dinnsen DA, Elbert M. Impact of substitution patterns on phonological learning by misarticulating children. Clinical Linguistics & Phonetics. 1997; 11(1):63–76.

Gross GH, St. Louis KO, Ruscello D, Hull FM. Language abilities of articulatory-disordered school children with multiple or residual errors. Language, Speech and Hearing Services in Schools. 1985; 16:171–186.

Hall PK. The occurrence of developmental apraxia of speech in a mild articulation disorder: A case study. Journal of Communication Disorders. 1989; 22:265–276. [PubMed: 2794108]

Hillenbrand J, Getty LA, Clark MJ, Wheeler K. Acoustical characteristics of American English vowels. Journal of the Acoustical Society of America. 1995; 97(5):3099–3111. [PubMed: 7759650]

Holm A, Crosbie S, Dodd B. Differentiating normal variability from inconsistency in children's speech: normative data. International Journal of Language & Communication Disorders. 2007; 42(4):467–486. [PubMed: 17613100]

Irwin JV, Huskey R, Knight N, Oltman S. A longitudinal study of the spontaneous remission of articulatory defects of 1665 school children in grades 1, 2, and 3: II. The sample. Acta Symbolica. 1974; 5(2):1–7.

Irwin JV, Knight N, Oltman S. A longitudinal study of the spontaneous remission of articulatory defects of 1665 school children in grades 1, 2, and 3: III. The study group. Acta Symbolica. 1974; 5(2):9–17.

Kent RD, Forner LL. Developmental study of vowel formant frequencies in an imitation task. Journal of the Acoustical Society of America. 1979; 65:208–217. [PubMed: 422815]

Kent R, Forner L. Speech segment duration in sentence recitations by children and adults. Journal of Phonetics. 1980; 8(2):157–168.

Kent RD, Kent JF, Duffy JR, Thomas JE, Weismer G, Stuntebeck S. Ataxic Dysarthria. Journal of Speech and Hearing Research. 2000; 43(5):1275–1289.

Kent RD, Kim YJ. Toward an acoustic typology of motor speech disorders. Clinical Linguistics & Phonetics. 2003; 17(6):427–445. [PubMed: 14564830]

Koenig LL. Distributional characteristics of VOT in children's voiceless aspirated stops and interpretation of developmental trends. Journal of Speech, Language, and Hearing Research. 2001; 44:1058–1068.

Koenig LL, Lucero JC, Perlman E. Speech production variability in fricatives of children and adults: Results of functional data analysis. The Journal of the Acoustical Society of America. 2008; 124(5):3158–3170. [PubMed: 19045800]

Lee S, Potamianos A, Narayanan S. Acoustics of children's speech: Developmental changes of temporal and spectral parameters. Journal of the Acoustical Society of America. 1999; 105:1455–1468. [PubMed: 10089598]

Lewis BA, Freebairn LA, Hansen AJ, Iyengar SK, Taylor HG. School-age follow-up of children with childhood apraxia of speech. Language, Speech, and Hearing Services in Schools. 2004; 35:122–140.

Lisker L, Abramson AS. A cross-language study of voicing in initial stops: Acoustical measurements. Word. 1964; 20:384–422.

Marquardt TP, Jacks A, Davis BL. Token-to-token variability in developmental apraxia of speech: three longitudinal case studies. Clinical Linguistics & Phonetics. 2004; 18(2):127–144. [PubMed: 15086134]

McNutt JC. Oral sensory and motor behaviors of children with /s/ or /r/ misarticulations. Journal of Speech and Hearing Research. 1977; 20:694–703. [PubMed: 604683]

Munson B. Variability in /s/ production in children and adults: Evidence from dynamic measures of spectral mean. Journal of Speech, Language, and Hearing Research. 2004; 47:58–69.

Nijland L, Maassen B, van der Meulen S. Evidence of motor programming deficits of children diagnosed with DAS. Journal of Speech, Language, and Hearing Research. 2003; 46:437–450.

Ohde RN. Fundamental frequency correlates of stop consonant voicing and vowel quality in the speech of preadolescent children. Journal of the Acoustical Society of America. 1985; 78:1554–1561. [PubMed: 4067069]

Pascoe, M.; Stackhouse, J.; Wells, B. Persisting speech difficulties in children. Wiley; Chichester, England: 2006.

Preston JL, Edwards J. Phonological processing skills of adolescents with residual speech sound errors. Language, Speech and Hearing Services in Schools. 2007; 38:297–308.

Preston JL, Edwards ML. Speed and accuracy of rapid speech output by adolescents with residual speech sound errors including rhotics. Clinical Linguistics & Phonetics. 2009; 23(4):301–318. [PubMed: 19382016]

Seeff-Gabriel, B.; Chiat, S.; Dodd, B. The relationship between speech disorders and language. In: Dodd, B., editor. Differential diagnosis and treatment of children with speech disorders. 2nd ed.. Whurr Publishers; London: 2005. p. 100-116.

Semel, E.; Wiig, EH.; Secord, WA. Clinical Evaluation of Language Fundamentals. 4 ed.. Harcourt Assessment; San Antonio, TX: 2003.

Sharkey SG, Folkins JW. Variability in lip and jaw movements in children and adults: Implications for the development of speech motor control. Journal of Speech and Hearing Research. 1985; 28:8–15. [PubMed: 3982001]

Shriberg LD. Five subtypes of developmental phonological disorders. Clinics in Communication Disorders. 1994; 4(1):38–53. [PubMed: 8019550]

Shriberg, LD. Childhood speech sound disorders: From postbehaviorism to the postgenomic era. In: Paul, R.; Flipsen, P., editors. Speech Sound Disorders in Children. Plural Publishing; San Diego: 2009.

Shriberg LD, Austin D, Lewis BA, McSweeny JL, Wilson DL. The speech disorders classification system (SDCS): extensions and lifespan reference data. Journal of Speech, Language & Hearing Research. 1997a; 40(4):723–740.

Shriberg LD, Austin D, Lewis BA, McSweeny JL, Wilson DL. The percentage of consonants correct (PCC) metric: Extensions and reliability data. Journal of Speech, Language, and Hearing Research. 1997b; 40:708–722.

Shriberg LD, Gruber FA, Kwiatkowski J. Developmental phonological disorders III: Long-term speech-sound normalization. Journal of Speech and Hearing Research. 1994; 37:1151–1177. [PubMed: 7823558]

Shriberg LD, Kwiatkowski J. Phonological disorders III: A procedure for assessing severity of involvement. Journal of Speech and Hearing Disorders. 1982; 47:242–256. [PubMed: 7186560]

Shriberg LD, Kwiatkowski J, Hoffmann K. A procedure for phonetic transcriptions by consensus. Journal of Speech & Hearing Research. 1984; 27(3):456–465. [PubMed: 6482415]

Shriberg LD, Lof GL. Reliability studies in broad and narrow phonetic transcription. Clinical Linguistics & Phonetics. 1991; 5(3):225–279.

Shriner T, Holloway M, Daniloff R. The relationship between articulatory deficits and syntax in speech defective children. Journal of Speech and Hearing Research. 1969; 12:319–325. [PubMed: 5808858]

Shuster LI, Ruscello DM, Haines KB. Acoustic patterns of an adolescent with multiple articulation errors. Journal of Communication Disorders. 1992; 25:165–174. [PubMed: 1487565]

Shuster LI, Wambaugh JL. Token-to-token variability in adult apraxia of speech: A perceptual analysis. Aphasiology. 2008; 22(6):655–669.

Smit AB, Hand L, Freilinger JJ, Bernthal JE, Bird A. The Iowa articulation norms project and its Nebraska replication. Journal of Speech and Hearing Disorders. 1990; 55(4): 779–797. [PubMed: 2232757]

Smith BL, McLean-Muse A. Articulatory movement characteristics of labial consonant productions by children and adults. Journal of the Acoustical Society of America. 1986; 80(5):1321–1328. [PubMed: 3782608]

Stathopoulos ET. Variability revisited: an acoustic, aerodynamic, and respiratory kinematic comparison of children and adults during speech. Journal of Phonetics. 1995; 23(1–2):67–80.

Sussman HM, Marquardt TP, Doyle J. An acoustic analysis of phonemic integrity and contrastiveness in developmental apraxia of speech. Journal of Medical Speech-Language Pathology. 2000; 8(4): 301–313.

Thoonen G, Maassen B, Gabreels F, Schreuder R, de Swart B. Towards a standardised assessment procedure for developmental apraxia of speech. European Journal of Disorders of Communication. 1997; 32(1):37–60. [PubMed: 9135712]

Thoonen G, Maassen B, Wit J, Gabreels F, Schreuder R. The integrated use of maximum performance tasks in differential diagnostic evaluations among children with motor speech disorders. Clinical Linguistics & Phonetics. 1996; 10(4):311–336.

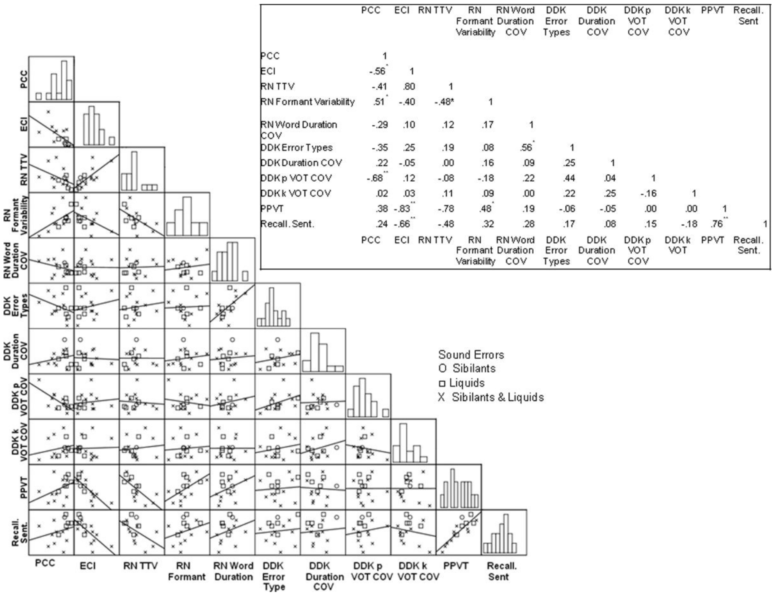
Tyler AA, Lewis KE. Relationships among consistency/variability and other phonological measures over time. Topics in Language Disorders. 2005; 25(3):243–253.

Tyler AA, Lewis KE, Welch CM. Predictors of phonological change following intervention. American Journal of Speech-Language Pathology. 2003; 12(3):289–298. [PubMed: 12971818]

Tyler AA, Williams M, Lewis K. Error consistency and the evaluation of treatment outcomes. Clinical Linguistics & Phonetics. 2006; 20(6):411–422. [PubMed: 16815788]

van Lieshout P, Merrick G, Goldstein LM. An articulatory phonology perspective on rhotic articulation problems: A descriptive case study. Asia Pacific Journal of Speech, Language, and Hearing. 2008; 11(4):283–303.

Walsh B, Smith A. Articulatory movements in adolescents: Evidence for protracted development of speech motor control processes. Journal of Speech, Language, and Hearing Research. 2002; 45:1119–1133.



	PCC	ECI	RN TTV	RN Formant Variability	RN Word Duration COV	DDK Error Types	DDK Duration COV	DDK p VOT COV	DDK k VOT COV	PPVT	Recall. Sent
PCC	1										
ECI	-.56*	1									
RN TTV	-.41	.80	1								
RN Formant Variability	.51*	-.40	-.48*	1							
RN Word Duration COV	-.29	.10	.12	.17	1						
DDK Error Types	-.35	.25	.19	.08	.56*	1					
DDK Duration COV	.22	-.05	.00	.16	.09	.25	1				
DDK p VOT COV	-.68**	.12	-.08	-.18	.22	.44	.04	1			
DDK k VOT COV	.02	.03	.11	.09	.00	.22	.25	-.16	1		
PPVT	.38	-.83**	-.78	.48*	.19	-.06	-.05	.00	.00	1	
Recall. Sent.	.24	-.66**	-.48	.32	.28	.17	.08	.15	-.18	.76**	1
	PCC	ECI	RN TTV	RN Formant Variability	RN Word Duration COV	DDK Error Types	DDK Duration COV	DDK p VOT COV	DDK k VOT	PPVT	Recall. Sent.

Figure 1.

Histograms and scatterplots of measures of phonetic accuracy, phonetic variability and language.

Notes: Lines represent slopes (correlation coefficients). PCC= Percent Consonants Correct; ECI = Error Consistency index; RN= Rapid Naming task; TTV = Total Token Variability; COV = Coefficient of Variation; DDK = Diadochokinetic task; VOT = Voice onset time; PPVT = Peabody Picture Vocabulary Test; Recall Sent = Recalling Sentences subtest of the CELF-4.

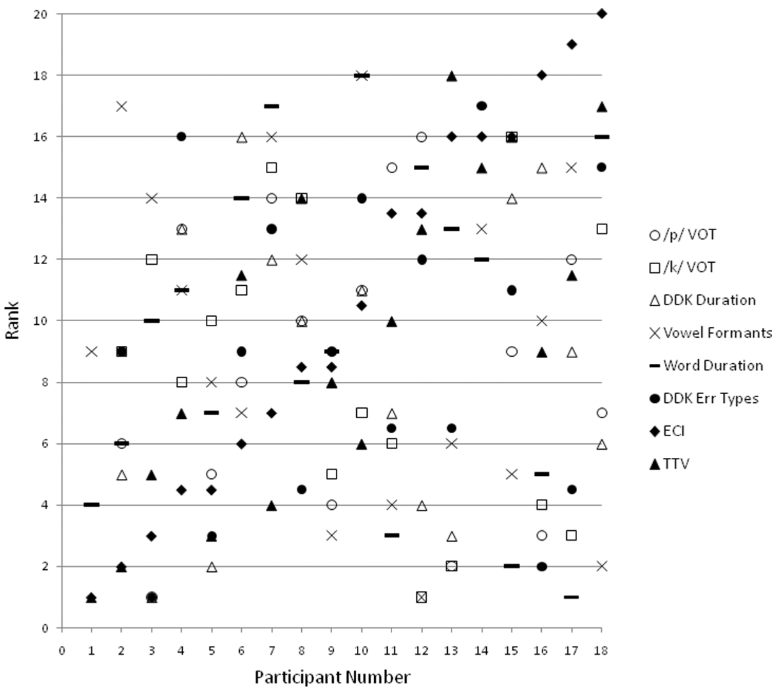


Figure 2.

Individual participants' ranks on measures of phonetic variability

Notes: Participant order (x-axis) arbitrarily ordered by ECI ranks. The y-axis values rank the participants ordinally according to their scores on each measure. Participants 1 and 14 did not complete the DDK task, and DDK acoustic measures were not collected from participant 7 due to a noisy signal.

Table 1

Tasks and phonetic variability measures

Task and Number of Speakers	**Transcription-based measures**	**Acoustic measures**
64 item picture naming task (n=20)	Percent accuracy: sibilants and liquids Error Consistency Index	–
Rapid multisyllabic picture naming (n=18)	Total Token Variability (mean of 6 words)	Formants (F1 and F2): Euclidean distances from the mean (average over 8 target vowels) Word durations: Coefficient of variation (mean of 6 words)
Diadochokinetic task (n=16)	Number of different realizations of /pʌtʌkʌ/in 40 attempts	Coefficient of Variation of VOT and trisyllable durations

Table 2A

Word duration criteria

Word	Onset event	Offset event
Elephant	Voicing onset	Offset of modal voicing for /n/.
Umbrella	Voicing onset	Offset of voicing [*] OR onset of noise for a following /s/, whichever came first.
Strawberries	Onset of /s/ noise	Offset of /z/ noise.
Thermometer	Voicing onset for /ɚ/	Offset of voicing [*] OR onset of noise for following /h/, whichever came first.
Helicopter	Voicing onset (which could co-occur with /h/ noise)	Offset of voicing [*] OR onset of noise for a following /s/, whichever came first.
Spaghetti	Onset of /s/ noise	Voicing offset [*]

[*] Where a word terminated in breath noise, the duration measures were made up to the point where the formants diverged from their vowel-like values.

Table 2B

Location of formant measurements

Word	**Vowel**	**Location**
Elephant	**/ɛ/**	Immediately after voicing stabilized.
Umbrella	**/ɛ/**	Maximum F2. This was the most reliable way to select a location with minimal effects of the two surrounding liquids.
Strawberries	/ɔ/	Approximate temporal midpoint between voicing onset after /t/ and closure for /b/.
Thermometer	/**a**/	Approximate temporal midpoint between preceding /m/ release and following /m/ closure.
Helicopter	/ɛ/	Immediately after formants stabilized. (Abduction and noise for /h/ may yield unstable formant tracking.)
Helicopter	/**a**/	Approximate temporal midpoint between voicing onset after /k/ and closure for /p/.
Spaghetti	/ɛ/	Approximate temporal midpoint between voicing onset after /**g**/ and /t/ closure. Where /t/ was realized as an approximant, the amplitude and formant traces were used to establish the transition between the vowels.
Spaghetti	/**i**/	Late in the vowel, when the F1–F2 difference was most extreme, and F2 and F3 were resolved. The /i/ values were intended to provide an extreme "anchor" in vowel space.

Table 3

Reliability data for acoustic measures: *r*-values, mean differences, and standard deviations (SDs) between original and re–measured data

Intra-rater reliability				
Task	Measure	*r*	Mean difference	SD of difference
RN	Word duration	.993	1 ms	20 ms
RN	Fl	.946	7 Hz	53 Hz
RN	F2	.978	22 Hz	85 Hz
DDK	VOT	.941	< 1 ms	7 ms
DDK	Trisyllable durations	.992	3 ms	16 ms

Inter-rater reliability				
Task	Measure	*r*	Mean difference	SD of difference
RN	Word duration	.874	6 ms	63 ms
RN	F1	.876	5 Hz	47 Hz
RN	F2	.957	11 Hz	67 Hz
DDK	VOT	.967	< 1ms	5 ms
DDK	Trisyllable durations	.994	1 ms	15 ms

Note: All correlations significant at $p < 0.01$.

Table 4

Descriptive statistics

Task	Measure	Mean	SD	Range
Picture Naming	Percent Consonants Correct	78.7	7.7	58–89
Picture Naming	Error Consistency Index	17.8	6.8	5–36
RN	Total Token Variability	.26	0.11	0.10–0.46
RN	Vowel F1, F2 Euclidean Distances from the mean (Hz)	87	23	44–139
RN	Word Duration COV	13	4	7–21
DDK	No. different error forms	7.9	4.2	1–17
DDK	Trisyllable duration (msec)	421	74	315–583
DDK	Trisyllable duration COV	15	5	8–26
DDK	/p/ mean VOT (msec)	41	9	25–55
DDK	/p/ VOT COV	39	14	20–78
DDK	/k/ mean VOT (msec)	35	14	12–55
DDK	/k/ VOT COV	30	12	16–55
PPVT–III	Standard Score	103	15	76–130
Recalling Sentences	Scaled Score	8.5	3.3	1–14

Notes: RN= Rapid Naming; COV=Coefficient of Variation; DDK = Diadochokinetic task; VOT= Voice Onset Time; PPVT–III = Peabody Picture Vocabulary Test–III.