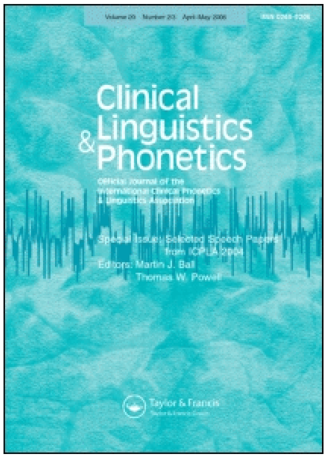

Clinical Linguistics & Phonetics
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713693308

Token-to-token variability in developmental apraxia of speech: three longitudinal case studies

Thomas P. Marquardt [a]; Adam Jacks [a]; Barbara L. Davis [a]
[a] Department of Communication Sciences and Disorders, The University of Texas, Austin, TX 78 712, USA

Online Publication Date: 01 March 2004
To cite this Article: Marquardt, Thomas P., Jacks, Adam and Davis, Barbara L. (2004) 'Token-to-token variability in developmental apraxia of speech: three longitudinal case studies', Clinical Linguistics & Phonetics, 18:2, 127 - 144
To link to this article: DOI: 10.1080/02699200310001615050
URL: http://dx.doi.org/10.1080/02699200310001615050

CLINICAL LINGUISTICS & PHONETICS, 2004, VOL. 18, NO. 2, 127–144 Taylor & Francis healthsciences

Token-to-token variability in developmental apraxia of speech: three longitudinal case studies

THOMAS P. MARQUARDT, ADAM JACKS and BARBARA L. DAVIS

Department of Communication Sciences and Disorders, The University of Texas, Austin, TX 78 712, USA

(*Received 31 January 2003; accepted 6 August 2003*)

Abstract

Variability in the speech production patterns of children with developmental apraxia of speech (DAS) was investigated in a three-year longitudinal study of three children with DAS. A metric was developed to measure token-to-token variability in repeated word productions from connected speech samples. Results suggest that high levels of total token and error token variability and low levels of word target stability and token accuracy characterize the disorder. Overall levels of variability and patterns of change over time differed between participants. Longitudinal patterns were indicative of decreasing total token variability and increasing token accuracy. However, change was not consistently unidirectional for two of the three children in the study, suggesting day-to-day performance differences in addition to within-session variability.

Keywords: Variability, apraxia of speech, articulation disorder, developmental speech disorder.

Introduction

Developmental apraxia of speech (DAS) is a controversial disorder marked by impaired ability to program speech movements in the absence of neuromuscular deficits. The disorder presumably arises from neurological injury not consistently documented from neuroimaging (e.g. Horwitz, 1984) or, more likely, from a failure in neural maturation/organization in cortical zones critical for phonemic representation and motor programming (see Marquardt, Sussman and Davis, 2001 for a review).

Address correspondence to: Thomas P. Marquardt, Department of Communication Sciences and Disorders, The University of Texas at Austin, 1 University Station, A1100, Austin, TX 78 712, USA. e-mail: tmarq@mail.utexas.edu

Clinical Linguistics & Phonetics
ISSN 0269-9206 print/ISSN 1464-5076 online © 2004 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/02699200310001615050

Diagnosis of DAS typically is on the basis of a group of behavioural features (e.g. vowel errors, prosodic abnormalities), but there is no clear consensus regarding which characteristics are essential to identification of the disorder and criteria for the selection of research participants have not been consistently applied (Marquardt *et al.*, 2001). The failure to reach consensus has yielded differing theoretical perspectives for DAS. Stackhouse and Snowling (1992) recommended use of the term 'developmental verbal dyspraxia' to emphasize non-speech specific aspects of the disorder such as reading and written language deficits. Square (1994) and Robin (1992) argued that the disorder may occur in isolation independent from associated language processes. Crary (1993), in a broader view, proposed that DAS is best considered as part of a continuum of impaired motor and language abilities. Clearly, there is a need to establish an agreed upon behavioural profile for DAS to facilitate selection of participants for research, to provide reliable criteria for differential diagnosis, and to assess the efficacy of treatment intervention paradigms.

Davis, Jakielski and Marquardt (1998) identified a cluster of speech characteristics for differential diagnosis of DAS that can be employed to identify children with the disorder (Skinder, Strand and Mignerey, 1998). Three speech characteristics (prosodic abnormalities, vowel errors, error variability) were identified by Davis *et al.* (1998) as potentially salient behavioural markers for the disorder. In this study we investigated one of these characteristics, speech production variability, in children with DAS.

Children with DAS are 'inconsistent' or 'variable' in the production of speech sounds (e.g. Morley, 1957; Rosenbek and Wertz, 1972; Ferry, Hall and Hicks, 1975; Aram and Glasson, 1979; Willams, Ingham and Rosenthal, 1981; Murdoch, Porter, Younger and Ozanne, 1984). Experimental evidence for increased variability has been equivocal, however, because of the lack of operational definitions for consistency and variability (Guyette and Diedrich, 1981; Hall, Jordan and Robin, 1993). Variability may be defined as repeated productions (i.e. word, phonetic sequences) that are different in the absence of contextual variation (Miller, 1992). As such, variability may be attributed to instability of the neural processes responsible for the programming and execution of phonetic sequences. Defined in this manner, several experimental studies of DAS provide evidence for increased variability in speech production.

Smith, Marquardt, Cannito and Davis (1994), in a study of repeated productions of vowels in isolation and in words, reported increased variability for a DAS child for first (F1) and second (F2) formant frequencies and for vowel duration. Mean coefficients of variation for F1 and F2 were nearly twice as great for a child with DAS than for typically developing and phonologically delayed children. Sussman, Marquardt and Doyle (2000) studied coarticulation via locus equations in DAS and normally developing children. For adult speakers and normal children, the frequencies for the F2 transition and F2 vowel cluster tightly around a line described by a linear regression function (e.g. the locus equation). Consonant placement, as described by F2 transition values, can be predicted on the basis of the F2 value of the following vowel. Children with DAS demonstrated increased variability in consonant placement, as described by locus equations, with values for mean standard error of estimate greater than those for typically developing children.

Phonetic variability also has been investigated in single word and connected speech tasks. In a comparative study of children with DAS, functional articulation disorder (FAD), and typical development, Schumacher, McNeil, Vetter and Yoder (1986) examined whole-token phonetic variability from elicited repetitions of words. Variability and consistency measures were based on similarity of three consecutively produced words, regardless of accuracy, using a qualitative rating scale. They found that a measure of total consistency comprised of measures of variability and consistency successfully distinguished children with DAS and FAD, with DAS children demonstrating lower total consistency.

Shriberg, Aram and Kwiatkowski (1997) analysed consistency of error types from repeated productions of connected speech samples in children with 'suspected' developmental apraxia of speech. Shriberg *et al.* (1997) computed a ratio of most common error class produced per lexical item to the total number of items produced, yielding a measure of error consistency. They found that error consistency did not distinguish children with suspected DAS from those with other speech disorders. However, selection of participants was based on referral with DAS and not an established set of diagnostic criteria so that direct comparison to other studies is not possible.

King, Jakielski and Malone (2001) explored variability in repeated productions from connected speech samples in a 22-month longitudinal case study of a child with DAS. They examined whole-token phonetic variability for both correct and incorrect attempts at words produced in three or more different recording sessions. Of 52 lexical types produced, only two were produced consistently, with the remaining 50 types varying in token production. Among the 50 unstable lexical types, they found that the child produced an average of 5.3 variant tokens per lexical type.

Children's speech movements are more variable than those of adults. This variability may be due to incomplete maturation of the neuromotor system or to the functional problem of meeting perceptual targets optimized for adult speakers while using child-sized articulators. Studies of phonological patterns in pre-kindergarten children (Gordon-Brannan and Hodson, 2000) suggest that accuracy of production is well-developed for many children by 4 to 5 years of age. Based on kinematic and acoustic studies, children develop adult-like stability for most articulatory parameters by age six or seven (Goffman and Smith, 1999), with jaw movement stabilizing earlier than lip movement (Sharkey and Folkins, 1985; Green, Moore, Higashikawa and Steeve, 2000; Green, Moore and Reilly, 2002). However, instability increases with increased linguistic or phonetic complexity of the speaking task (Sussman, Duder, Dalston and Cacciatore, 1999) and may persist until children are as old as 11 years (Smith, 1994).

Increased variability may be expected in the speech production of typically developing children with adult-like stability attained by the age of 7. Empirical studies examining speech production variability in DAS, although limited in scope and diverse in methodology, suggest that variability may be a critical feature of the disorder. The purpose of this study was to investigate longitudinal changes in phonetic variability in connected speech samples for three children with developmental apraxia of speech.

Method

Participants

Three English-speaking male children (P1, P2 and P3) referred for differential diagnosis of developmental apraxia of speech (DAS) participated in the study. Diagnosis of DAS was based on identification of a cluster of speech and language characteristics typical of the disorder, including prosodic abnormalities, vowel errors, high frequency of consonant and syllable omissions, and segmental variability (Davis *et al.*, 1998). When available, previous assessments of hearing, cognition, and expressive and receptive language were used to establish participant status. Participants demonstrated auditory acuity within normal limits bilaterally based on pure tone audiometric screening at octave frequencies from 250 Hz to 4000 Hz. Review of clinical records revealed no positive history for emotional disorders or fluctuations in auditory acuity due to middle ear disorders. An oral mechanism examination, diadochokinetic tasks, and informal oral and limb praxis tasks revealed no obvious signs of structural abnormalities or dysarthria.

The participants were assessed at one-year intervals (Time 1, Time 2 and Time 3) between the ages of 4;6 and 7;7 to track longitudinal changes in speech production. The assessment included articulation testing and spontaneous speech sampling.

Participant descriptions

Speech profiles for participants are shown in table 1, including age at each assessment, scores on single-word articulation tests, and accuracy of consonant and vowel production from connected speech samples. Relational analyses of consonant

Table 1. *Participant data: age, articulation test results, and accuracy of consonant and vowel production at three times*

		Time 1	Time 2	Time 3
P1	Age	4;6	5;5	6;5
	Articulation Test	Profound[1]	3rd %ile[2]	31 %ile[3]
	Consonant Accuracy	71	61	76
	Vowel Accuracy	61	72	85
P2	Age	5;10	6;10	7;7
	Articulation Test	Severe[1]	4 %ile[2]	16 %ile[2]
	Consonant Accuracy	61	72	69
	Vowel Accuracy	75	85	71
P3	Age	5;6	6;10	7;5
	Articulation Test	<1 %ile[2]	<1 %ile[2]	<1 %ile[2]
	Consonant Accuracy	34	63	74
	Vowel Accuracy	65	70	76

Note: Consonant and vowel accuracy represent relational measures derived from analysis of entire connected speech sample from each time point.
[1] *Assessment of Phonological Processes-Revised* (APP-R, Hodson, 1986).
[2] *Goldman-Fristoe Test of Articulation* (GFTA, Goldman and Fristoe, 1986).
[3] *Khan-Lewis Phonological Assessment* (KLPA, Khan and Lewis, 1986).

and vowel production were based on connected speech samples collected at each time point which also served as the source of data for variability analyses.

Participant P1

P1 was evaluated at the ages of 4;6, 5;5 and 6;5 years. Developmental history was within normal limits with the exception of speech and language. First words were produced at 22 months. Speech and language evaluation was completed at 3;5 with unscored administration of the *Goldman-Fristoe Test of Articulation* (GFTA, Goldman and Fristoe, 1986) revealing a phonetic inventory of /p, n, w, b, f, d, t, ʃ, tʃ, l, θ, m, and dʒ/. Most productions consisted of single sounds or syllables. Administration of the *Preschool Language Scale-Revised* (PLS-R, Zimmerman, Steiner and Pond, 1979) showed auditory comprehension within normal limits (age equivalent = 3;7.5) with delayed verbal ability (age equivalent = 1;7.5). A second speech and language evaluation at age 3;8 found normal receptive language skills, a limited number of consistent word forms, restricted phonetic repertoire, rudimentary syllable shapes, and reduced ability to imitate nonverbal movements of the speech mechanism. Results from the *Test of Language Development-Primary: Second Edition* (TOLD-P:II, Newcomer and Hammill, 1988) completed at 4;1 suggested average single word comprehension (75^{th} percentile), with low average comprehension for word classes and grammatical morphemes and below average comprehension for elaborated sentences based on administration of the *Test of Auditory Comprehension of Language- Revised* (TACL-R, Carrow-Woolfolk, 1985). Neurological evaluation revealed mild dysarthria and oral apraxia. Follow-up comprehension and neurological evaluations at 5;0 were consistent with the earlier assessments.

Treatment was initiated at age 2;2 and continued for the total duration of the study. Initially treatment focused on increasing the accuracy of verbal communication with supplementary use of sign and communication boards to facilitate language and cognitive development. Subsequently, goals were to expand the phonemic repertoire, stabilize consistent sound production, and increase the use of more complex syllable shapes in words and phrases.

Single-word articulation improved over the course of the study, beginning with a profound rating on the *Assessment of Phonological Processes-Revised* (APP-R, Hodson, 1986), with percentile rankings of 3 and 31 on the GFTA and *Khan-Lewis Phonological Assessment* (KLPA, Khan and Lewis, 1986), respectively, at subsequent reassessments (see table 1). Improvements also were noted in relational analyses of connected speech samples, although the pattern of improvement was not uniform across measures. At Time 1, consonant production was 71% accurate, decreasing to 61% at Time 2, and increasing again to 76% at Time 3. Pl's vowel accuracy improved consistently across time (Time 1: 61%, Time 2: 72%, Time 3: 85%).

Participant P2

P2 had a normal developmental history with the exception of speech and language. At 2;6 results of the *Sequenced Inventory of Communication Development* (SICD, Hedrick, Prather and Tobin, 1984) revealed a profound expressive communication deficit (no basal score obtained) and a severe receptive deficit (receptive communication age = 1;4). Vocalizations were comprised exclusively of vowels. Administration of the *Developmental Profile II* (DPII, Alpern, Boll and Shearer, 1986) indicated age scores within normal range for subtests of physical, self-help, social, and academic skills (26–34 months), while communication age was 10 months. Results of the *Peabody Picture Vocabulary Test-Revised* (PPVT-R; Dunn and Dunn, 1981) administered at 5;0 suggested normal receptive language (53rd percentile). Administration of the *Blakeley Screening Test for Developmental Apraxia of Speech* (Blakeley, 1980) indicated a probability of greater than 99% of correct diagnosis of DAS. Results of the APP-R (Hodson, 1986) revealed frequent vowel errors, final consonant deletion, errors on initial consonants, glottal stops, and reduction of consonant clusters. Following this evaluation, P2 returned for follow-up evaluations at approximately yearly intervals (5;10, 6;10, 7;7 years of age).

P2 began therapeutic intervention at 2;6 and had completed 5 years of treatment by the end of the study. Treatment goals focused on expanding the phonemic repertoire, increasing the accuracy of frequently incorrect consonants and vowels, and increasing the use of consonant clusters. Gains in articulation proficiency were apparent as measured by single-word articulation tests (table 1), with a severe rating on the APP-R at Time 1 and percentile rankings of 4 and 16 on the GFTA at Time 2 and Time 3, respectively.

Measures of consonant and vowel accuracy from connected speech samples indicated general improvement over time, with better consonant production at Time 3 than Time 1 although with best performance at Time 2 (Time 1: 61%, Time 2: 72%, Time 3: 69%). However, vowel production, while highest in accuracy at Time 2, was least accurate at Time 3 (Time 1: 75%, Time 2: 85%, Time 3: 71%). While measures of segmental accuracy were the greatest at Time 2, these levels are indicative of significant impairment of speech production even at best performance.

Participant P3

P3 had a normal developmental history with the exception of speech and language, sitting unsupported and walking before 1 year of age. He began saying words between 1 and 2 years, although his mother reported that productions were 'unclear'. P3 was evaluated for communication deficits at 4;2 at his elementary school. Administration of the APP-R (Hodson, 1986) indicated a profound level of disorder, with overall average percentage of phonological process use of 83%. At that time, the only consonants in his production repertoire were /h/ and /n/. P3 was first referred at 5;7 by his school speech-language pathologist for differential diagnosis of DAS. Analysis of a spontaneous speech and language sample revealed an incomplete phonetic inventory, variable consonant and vowel errors, unusual and persistent phonological processes, inconsistency of stress and intonation, and nasality. Administration of the PPVT-R (Dunn and Dunn, 1981) at 5;6 indicated comprehension of single words within normal limits, with a percentile rank of 32. On the basis of this evaluation and previous speech and language assessments, P3 was diagnosed with DAS. P3 returned at 5;11 for testing of short-term auditory memory and metalinguistic skills. He demonstrated difficulty in repeating more than two lexical items, identifying words with more than two syllables, and identifying CVC minimal word pairs. He was seen for yearly evaluations at the ages of 5;6, 6;10, and 7;5 years.

P3 began receiving speech and language therapy at the age of 4;2 years with a focus on increasing the accuracy of vowel and consonant productions and increased syllabic complexity. In contrast to P1 and P2, P3 demonstrated negligible gains in single-word articulation testing, with percentile rankings on the GFTA of less than one at all three time points (see table 1). However, he made dramatic improvements in segmental production in connected speech. P3's consonant accuracy improved at each consecutive time point (Time 1: 34%, Time 2: 63%, Time 3: 74%). Vowel accuracy showed a similar pattern of improvement, although with less dramatic increases compared to consonant accuracy (Time 1: 65%, Time 2: 70%, Time 3: 76%). His patterns of improvement were more regular than P1 and P2, with generally continuous gains from Time 1 to Time 3. Nevertheless, segmental accuracy remained clearly deficient at Time 3.

Data collection

In each of the follow-up evaluations for this longitudinal study, connected speech and language samples were collected by a graduate clinician and a certified speech-language pathologist experienced in diagnosis of DAS, yielding a total of nine 1-hour samples (one sample at each of the three time points for the three participants). Age appropriate materials were employed for sample collection. Connected samples were assessed as representative of the child's speech patterns based on parent interviews at the time of evaluation.

Data analysis

Spontaneous samples were phonetically transcribed by a certified speech-language pathologist using broad phonetic transcription. Transcription reliability was completed for 10% of the nine samples with an average of 86.22% and a range of 75 to 96.26% agreement for all samples. Analyses of phonetic inventory and accuracy were completed for consonants, vowels, and word shapes related to length of utterance for each participant.

Repeated word tokens were identified from the speech samples, defined as any intended target word form occurring at least twice in the sample. Each token production was compared to the target word, and a series of analyses was used to investigate accuracy of tokens, stability of word targets and variability. Word tokens were compared to target forms, with each token judged as accurate when all phonetic segments in the target production were produced correctly. Allophonic variations of consonants (e.g. aspiration) were not considered errors, nor were small variations in vowel production (e.g. /i/→/ɪ/ substitution). For variability and stability measures, token productions for each target were compared to determine the number of different variants for a target, regardless of accuracy (e.g. the number of different ways the target was produced).

Table 2 shows the formulas used to calculate token *accuracy, target stability, total token variability* and *error token variability*. Examples of repeated word productions are included to illustrate the measures.

Token accuracy was defined as the percentage of word tokens matching the target compared to the total number of tokens produced. For example, if two of four tokens were correct, token accuracy is 50% for that target (see example 2 in table 2).

Table 2. *Sample repeated token data and corresponding measures of token accuracy, target stability, and token-to-token variability*

Example	Gloss	Target	Token 1	Token 2	Token 3	Token 4	Token accuracy[1]	Target stability[2]	Total token variability[3]	Error token variability[4]
1	‘cat’	/kæt/	/kæt/	/kɛt/	/kæt/	/kit/	2/4 = 50%	0	(3 variants −1 dF)/ (4 tokens− 1 dF) = 2/3 = 67%	(2 incorrect variants −1 dF)/ (2 incorrect tokens− 1 dF) = 1/1 = 100%
2	‘plate’	/pleɪt/	/peɪt/	/pɛt/	/pleɪ/		0/3 = 0%	0	(3 variants −1 dF)/ (3 tokens− 1 dF) = 2/2 = 100%	(3 incorrect variants −1 dF)/ (3 incorrect tokens− 1 dF) = 2/2 = 100%
3	‘dog’	/dag/	/dag/	/dag/	/dag/	/dag/	4/4 = 100%	1	(1 variant−1 dF)/ (4 tokens −1 dF) = 0/3 = 0%	N/A- no incorrect tokens
4	‘chair’	/tʃeɪr/	/teɪr/	/teɪr/	/teɪr/		0/3 = 0%	1	(1 variant−1 dF)/ (3 tokens−1 dF) = 0/2 = 0%	(1 incorrect variant−1 dF)/ (3 incorrect tokens−1 dF) = 0/2 = 0%

[1]Token accuracy = (# tokens matching the target)/(number of tokens produced).
[2]Target stability = 1 if all tokens are identical for a given target; 0 if any tokens are produced differently.
[3]Total token variability = (# variant productions−1 degree of freedom (dF))/(# tokens produced−1 dF).
[4]Error token variability = (# incorrect variant productions−1 degree of freedom (dF))/(# incorrect tokens produced−1 dF).

Target stability refers to percentage of word targets for which all repeated tokens were produced alike, regardless of correctness. In table 2, two of the four targets were stable (examples 3 and 4); each token of 'dog' was produced the same (correctly), while each token of 'chair' was produced the same way, but incorrectly. For the whole sample listed, overall target stability was 50% (2 stable targets/4 total targets).

Total token variability (i.e. variability in all produced tokens) was assessed by computing a ratio of number of different variant productions to total number of tokens produced.[1] In example 1, the word target 'cat' was produced in three different ways, yielding a total variability of (3 variants − 1 df)/(4 tokens − 1 df) = 2/3, or 67%. Examples 3 and 4 represent word targets wherein each token is produced the same each time (regardless of accuracy), yielding total token variability of 0%, while example 2 is produced differently each time, with a corresponding 100% variability.

Error token variability (i.e. variability among incorrectly produced tokens) was assessed by computing a ratio of number of different types of incorrect productions to total number of incorrect tokens produced.[2] In examples 1 and 2 from table 2, error token variability is 100% because all of the incorrectly produced tokens were produced differently. In example 4, error token variability is 0% because all errors were identically incorrect. No error token variability was calculated for targets that did not include any incorrect productions; hence for example 3 this measure is not applicable.

Results

The analyses of repeated tokens yielded measures of target stability, total token variability, error token variability, and token accuracy for the three participants in the study. Results are presented as three case studies to facilitate description of variations in performance with maturation of the participants.

Participant P1

P1 produced a total of 104 repeated word targets (Time 1: 25, Time 2: 33, Time 3: 46 targets) and a total of 577 repeated tokens (Time 1: 194, Time 2: 157, Time 3: 226 tokens). Results of the token analyses are shown in figure 1. Token accuracy decreased from Time 1 to Time 2, with a dramatic increase at Time 3 (Time 1: 47%, Time 2: 25%, Time 3: 64%). Target stability improved minimally from Time 1 to Time 2, with a large increase for Time 3 (Time 1: 20%, Time 2: 21%, Time 3: 44%). Total token variability increased from Time 1 to Time 2 and decreased at Time 3 (Time 1: 20%, Time 2: 36%, Time 3: 22%) while error token variability was greater than total variability, with consistent increases at the three measurement points (Time 1: 32%, Time 2: 40%, Time 3: 50%).

Patterns of change over time reveal an inverse relationship between token accuracy and total token variability, with variability decreasing as accuracy increases. This pattern is evident in both the interval between Time 1 and Time 2 (decreasing accuracy with increasing total token variability) and between Time 2 and Time 3 (increasing accuracy and decreasing total token variability). The inverse relationship between these two measures is logical; as the number of accurate productions increases, the number of variant productions must decrease.

P1 demonstrated moderate but consistently increasing error token variability. In

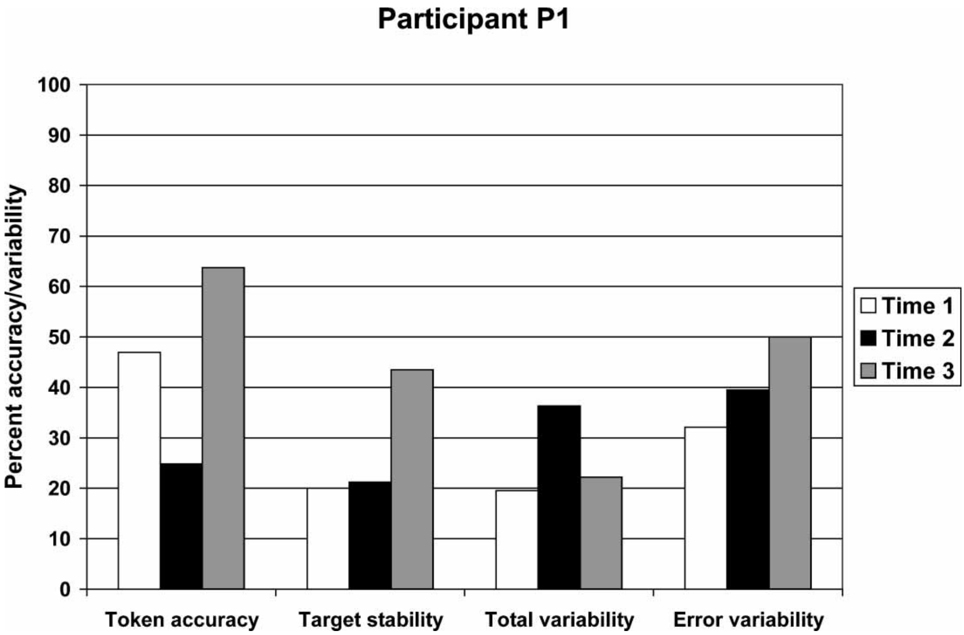

Figure 1. *Participant P1 token-to-token variability: percent token accuracy, target stability, total token variability, and error token variability at Time 1, Time 2, and Time 3.*

principle, increased error token variability can result from either increased number of incorrect variants or decreased total number of incorrect tokens. For P1 the pattern of increased error token variability resulted from small increases in the number of incorrect tokens with somewhat larger increases in incorrect variants (Time 1→Time 2) and a greater decrease in the number of incorrect tokens relative to the decrease in number of incorrect variants (Time 2→Time 3) (incorrect variants at Time 1: 48; Time 2: 66; Time 3: 56; incorrect tokens at Time 1: 103; Time 2: 118; Time 3: 82).

P1 demonstrated an unexpected pattern of change over time, as exemplified by the decrease in token accuracy from Time 1 to Time 2. Regardless of the direction of change, token accuracy and total token variability appeared to be inversely related. Error token variability increased due to the combination of increased number of incorrect variants (Time 1→Time 2) and decreased incorrect tokens (Time 2→Time 3). Target stability changed minimally from Time 1 to Time 2, and then doubled from Time 2 to Time 3. These results indicate a general pattern of improvement over time, although the lack of a consistent directional effect for accuracy and total token variability preclude a clear interpretation of this finding.

Participant P2

P2 produced a total of 121 repeated word targets, with two or more attempts at each target (Time 1: 34, Time 2: 54, Time 3: 33 targets). A total of 513 repeated tokens were produced for the three time points (Time 1: 103, Time 2: 257, Time 3: 153 tokens).

Results of token analyses for P2 are shown in figure 2. Token accuracy increased markedly from Time 1 to Time 2 and then decreased at Time 3 (Time 1: 33%, Time 2: 59%, Time 3: 46%). Target stability improved from Time 1 to Time 2, with a slight decrease at Time 3 (Time 1: 24%, Time 2: 43%, Time 3: 39%). Total token variability decreased from Time 1 to Time 2 and increased slightly at Time 3 (Time 1: 46%, Time 2: 19%, Time 3: 26%). Error token variability showed a similar pattern of initial decrease from Time 1 to Time 2 with a very slight increase at Time 3 (Time 1: 63%, Time 2: 33%, Time 3: 34%).

Individual patterns of change over time for accuracy and total variability were different than P1. However, the inverse relationship between token accuracy and total token variability was similar – as accuracy increased, total variability decreased (e.g. Time 1→Time 2), and vice versa (Time 2→Time 3).

Error token variability was highest at Time 1 and decreased at Time 2 and Time 3. Decreased error token variability from Time 1 to Time 2 appeared to be primarily related to a sharp increase in the number of incorrect tokens produced at Time 2 while number of different incorrect tokens changed very little (incorrect variants at Time 1: 53; Time 2: 59; Time 3: 45; incorrect productions at Time 1: 69; Time 2: 106; Time 3: 82). The increased number of incorrect tokens is not unexpected given that P1 produced more repeated tokens at Time 2 than at Time 1 and Time 3. However, the preponderance of repeated tokens at Time 2 is surprising given that repeated tokens came from speech samples of equal length (100 utterances). Nevertheless, it is notable that the number of incorrect variants remained constant over time, even as the total number of repeated tokens increased.

The results reflect a dramatic change from Time 1 to Time 2 for all measures,

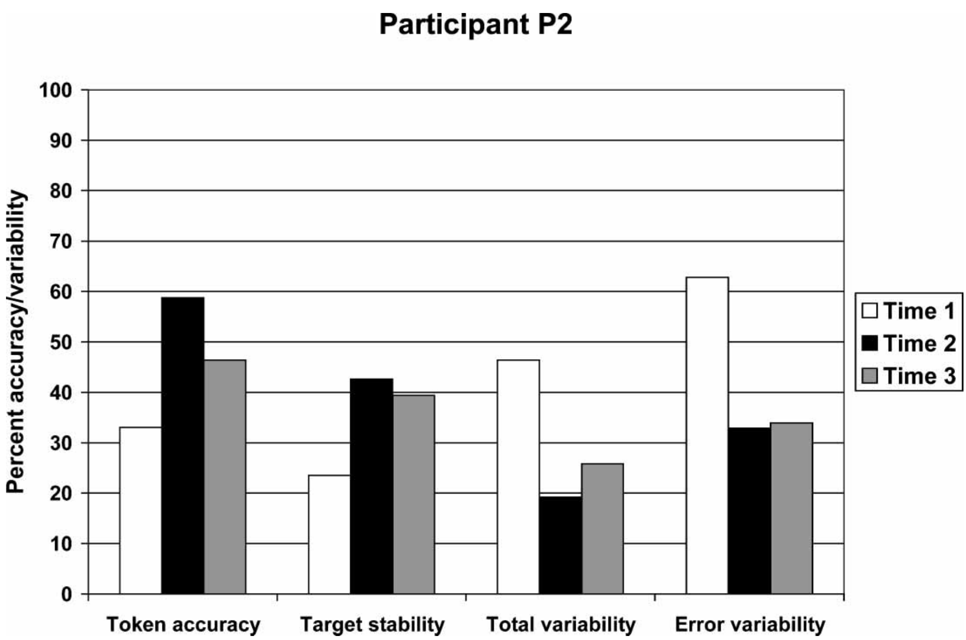

Figure 2. *Participant P2 token-to-token variability: percent token accuracy, target stability, total token variability, and error token variability at Time 1, Time 2, and Time 3.*

with higher token accuracy and target stability and lower total and error token variability. Most measures changed minimally between Time 2 and Time 3 (e.g. 1–7% for stability and variability), although a greater decrease (13%) was found in token accuracy. Token accuracy appeared to be inversely related to total token variability. Error token variability was affected primarily by fluctuations in number of incorrect tokens between time points, while number of incorrect variants remained constant. Results indicate general improvement over time, but decreased token accuracy from Time 2 to Time 3 suggests that day-to-day variations in speech production may be an important feature of DAS.

Participant P3

P3 produced a total of 131 repeated word targets, with two or more attempts at each target (Time 1: 33, Time 2: 57, Time 3: 41 targets). A total of 438 repeated tokens were produced across the three time points (Time 1: 115, Time 2: 188, Time 3: 135 tokens).

Results of the token analyses are shown in figure 3. Token accuracy increased consistently from Time 1 to Time 3 (Time 1: 12%, Time 2: 31%, Time 3: 47%). Target stability declined from Time 1 to Time 2 and increased at Time 3 (Time 1: 15%, Time 2: 12%, Time 3: 29%). Total token variability decreased minimally from Time 1 to Time 2 and more dramatically at Time 3 (Time 1: 62%, Time 2: 60%, Time 3: 43%). Error token variability varied minimally across the time points, with a small increase from Time 1 to Time 2 and a decrease at Time 3 (Time 1: 65%, Time 2: 72%, Time 3: 60%).

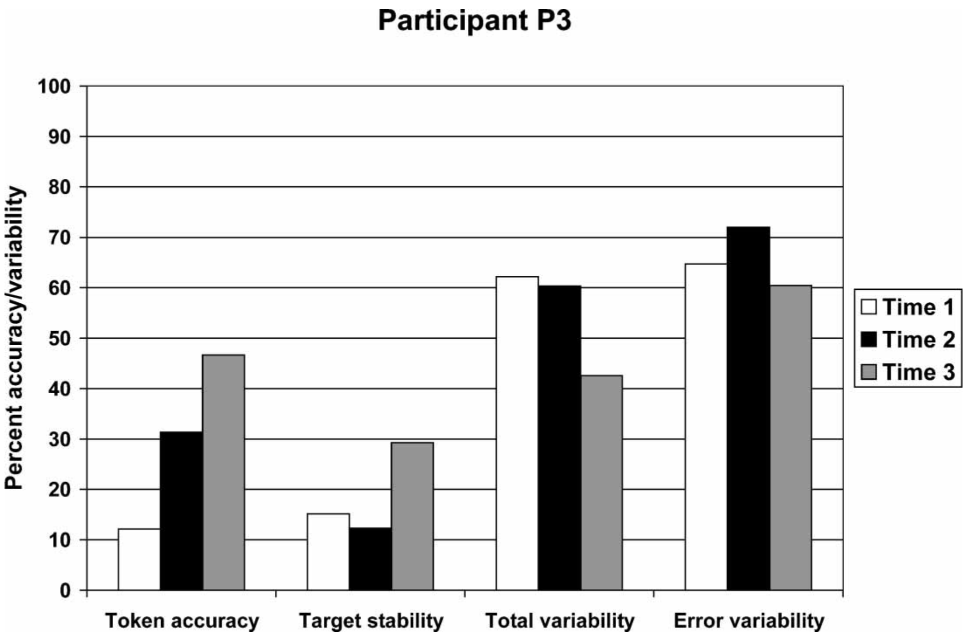

Figure 3. *Participant P3 token-to-token variability: percent token accuracy, target stability, total token variability, and error token variability at Time 1, Time 2, and Time 3.*

Total token variability for P3, like P2 and P3, decreased as token accuracy increased. In contrast, variability decreased from Time 1 to Time 2 and from Time 2 to Time 3. Error token variability varied minimally and was characterized by parallel patterns of change for number of incorrect variants and total number of incorrect tokens produced (number of incorrect variants at Time 1: 77; Time 2: 108; Time 3: 55; number of incorrect tokens at Time 1: 101; Time 2: 129; Time 3: 72).

P3 showed the most consistent pattern of change, although his speech sound development was notably more delayed than P1 and P2, with 47% token accuracy at maximum. Nevertheless, token accuracy increased at each time point while total token variability decreased. Target stability was very low at Time 1 and Time 2 but increased at Time 3. Error token variability was very high, between 60% and 72%, with very little change between time points.

In summary, no clear longitudinal patterns of change were evident in measures of variability in the connected speech of the three children with DAS in this study. Token accuracy improved for all participants, with higher accuracy at the end of the study. Target stability increased for all children, with only minor deviations. Total token variability decreased, with the exception of P1, who had very low total variability at Time 1. While the decreasing trend was not present at all intervals, a pattern was observed wherein total token variability was inversely related to token accuracy for all participants at all measurement points, with total token variability decreasing as token accuracy increased.

The direction of change was not unidirectional for any measure, particularly for P1 and P2. While P3 demonstrated the most consistent patterns of increased accuracy of production, he also was the most impaired of the three children, with token accuracy never exceeding 50%.

Error token variability, to a greater extent than other measures, showed a lack of a longitudinal pattern. Error token variability increased for P1, decreased for P2 and remained high for P3. For P3, both number of incorrect variants and incorrect tokens varied across time. In contrast, P1 and P2 varied minimally in the number of incorrect variants produced, and change in error token variability was related to variation in the number of incorrect tokens produced.

Group Performance

Low incidence disorders such as developmental apraxia of speech make it difficult to assemble large sample studies for inferential statistical analysis. In order to examine the relationships between measures of target stability and accuracy and measures of variability, and as a means to investigate potentially important parameters of the disorder for future study, linear regression best fit lines were determined for the pooled data from the three subjects at three measurement points (figure 4). Examination of the trends reflected in the best fit lines for total token variability suggest: (1) a highly positive relationship ($F(1,7)=25.96$, $p<0.01$, $r=0.89$, $R^2=0.79$) between error token variability and total token variability, and (2) negative relationships between total token variability and token accuracy ($F(1,7)=13.4$, $p<0.01$, $r=-0.81$, $R^2=0.66$) and stability ($F(1,7)=9.00$, $p<0.05$, $r=-0.75$, $R^2=0.56$). Clearly the data reflect a relationship of reduced variability with increases in token accuracy and stability.

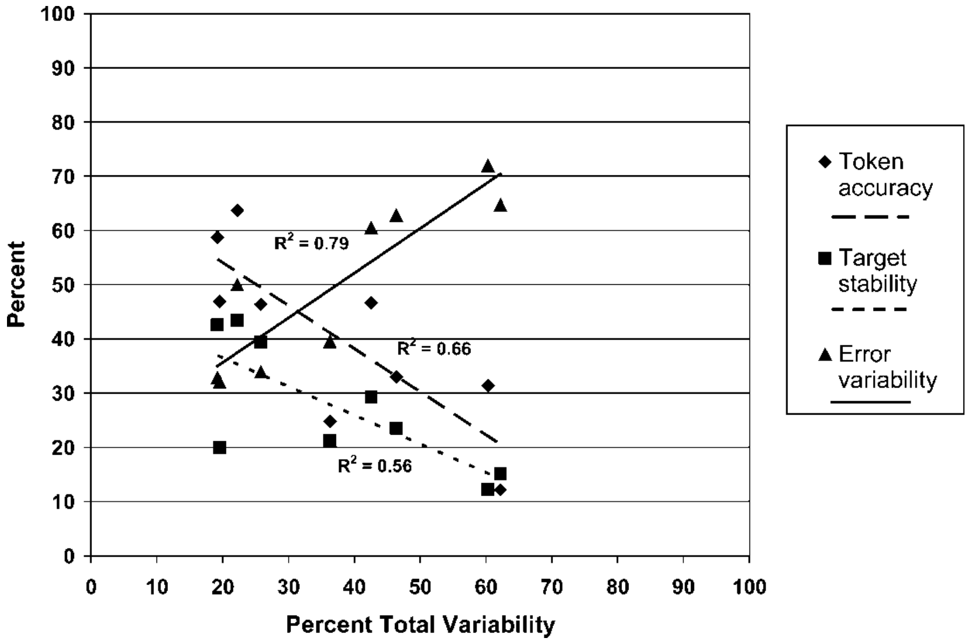

Figure 4. *Linear regression best fit lines for total token variability: token accuracy, target stability, and error token variability are regressed on total token variability, with data points pooled across time and participant.*

Discussion

Repeated tokens from connected speech samples were analyzed for accuracy of production, stability of word targets, and the diversity of variant productions in three children with developmental apraxia of speech. The children received speech therapy throughout the 3 year period of study. Measures of total token and error token variability were high in comparison to expected levels of functioning in typically developing children. A finding of highly variable productions of sounds and sequences of sounds is consistent with results from earlier studies of DAS (Schumacher *et al.*, 1986; Smith *et al.*, 1994) that found increased variability and reduced consistency of production. These results suggest that variability of motor performance may be integral to the disorder and an important marker for differential diagnosis.

Longitudinal trends for measures of variability were not consistent in the children in this study. Error token variability was least consistent; P3 maintained a high level of error token variability, P2 decreased and P1 increased in error token variability during the course of the study. While measures of stability were higher at Time 3 than Time 1 and total token variability was lower at Time 3 than Time 1 for P1 and P2, patterns of change for these measures was not consistent with predicted performance for all three participants at the three measurement points. These results suggest marked variability in session-to-session error productions in children with DAS. However, examination of pooled data for the three subjects revealed a trend for decreased variability with increased accuracy of production. Longitudinal study of a large DAS group might be expected to yield inconsistent sound and word productions and increases and decreases in variability and stability measures at contiguous time points for individual participants, with group performance characterized by improved accuracy and reduced variability.

Differences were noted between participants on single-word articulation testing and relational analyses of consonants and vowels in connected speech. These differences were paralleled in the variability results. P3 was most impaired, demonstrating higher levels of total and error token variability and lower levels of target stability and token accuracy than P1 and P2, with very few exceptions.

An expected finding in longitudinal studies of children with DAS (Stackhouse and Snowling, 1992) is increased segmental accuracy with maturation. P3 demonstrated improved consonant, vowel, and syllable shape accuracy during the study; the lowest level of accuracy was at Time 1 and the highest at Time 3. P1, however, declined in segmental and syllable shape accuracy from Time 1 to Time 2, with best performance at Time 3. P2 improved from Time 1 to Time 2 and then decreased from Time 2 to Time 3; his best performance was at Time 2.

The children with DAS received speech treatment throughout the 3 year period of the study, beginning at 2;2, 2;6 and 4;2 years of age for P1, P2, and P3, respectively. Therapeutic intervention was not systematically examined in this study, but treatment may have resulted in observed decreased variability and increased accuracy. Factors that may have affected the results included age of intervention and the focus of treatment. It is notable that P3, the child with the most profound deficits, began therapy latest, at 4;2 years. Treatment for all children focused primarily on improved segmental accuracy and increased syllabic complexity. Additionally, P1's treatment plan specifically targeted increasing stability for correct productions. His results showed increased target stability over the course of the study, although total token variability did not decrease consistently. It is likely that treatment of variable speech patterns had an effect on speech variability in the participants studied, but such a conclusion is speculative given the limited number of participants.

The finding of a high level of variability with a lack of a consistent pattern of improvement in speech accuracy has important clinical implications for the child with DAS. The presence of higher than expected variability from repeated assessment may be a key diagnostic marker and should be targeted within the speech and language evaluation. Due to large variations in performance, one-time measures of speech production skill from single word articulation tests and speech sample analyses may be insufficient in determining the presence and severity of the disorder. Differential diagnosis should rest on multiple assessments of articulation proficiency. Finally, treatment efficacy studies may require detailed analysis of a large corpus of speech production data rather than a limited number of probes to determine the effectiveness of behavioural intervention.

Direct comparisons of the variability findings of this study with previous studies of DAS (Shriberg *et al.*, 1997; Schumacher *et al.*, 1986) are limited because of differences in the measures employed. The 'variability' score used by Schumacher *et al.* assessed the likeness between elicited repetitions of tokens without respect to correctness, a measure similar to total token variability used in this study. The score differentiated children with DAS from those with functional articulation disorders at an alpha level of 0.01, with greatest variability for children with DAS in all three response conditions (elicited word, imitated word, imitated sentence). While

Schumacher *et al.*'s variability score was not identical to total token variability, the scores can be converted to allow a comparison of the two metrics. For each word target, Schumacher *et al.* counted a '0' if all three tokens were produced identically, a '1' if there were two variant productions, and a '2' if there were three variant productions for the three produced tokens. When converted to total token variability (TTV), a '0' represents 0% TTV, a '1' represents 50% TTV and a '2' represents 100% TTV. This is representative of a linear relationship expressed by the formula TTV = 50* variability. In the elicited words condition, Schumacher et al. reported a variability range from 0.5 to 0.9 for the DAS group, which corresponds to 25–45% TTV. In the imitated words condition, variability ranged from 0.3–1.0 (15–50% TTV) and in the imitated sentence condition variability ranged from about 0.3–1.2 (15–60% TTV). These values are similar to the range of 19–60% total token variability for the participants of this study.

Error token variability is closely related to the 'error consistency' reported by Shriberg *et al.* (1997). Shriberg *et al.* determined the consistency of specific error classes; the error token variability of this study investigated whole-token variants of incorrect productions. Assuming a complementary relationship between 'error consistency' and 'error variability', levels of 82–84% error consistency observed by Shriberg *et al.* for young children with suspected DAS corresponds to error variability of 16–18%. Error token variability was significantly higher (30–70%), which may be indicative of the more severe involvement of the DAS participants of this study.

Neuroimaging studies have not identified specific loci of neural dysfunction in DAS. Although speculative, we have argued that the disorder arises from a lack of the neural hard wiring required for phonemic representation that impacts both perceptual distinctiveness (Sussman *et al.*, 2000) and the motor algorithms that develop with perceptual categorization. Highly variable segmental production may be a manifestation of spatial targeting errors integral to the lack of neural instantiation of phonemic representation.

Notes

1. A simple ratio of variants to total tokens does not account for the fact that at least one variant (viz. the correct production) should occur for each word target attempted. Modifying the ratio by subtracting 1 degree of freedom from the numerator and the denominator allows the formula to account for 0% variability when only 1 variant (correct or incorrect) is produced (e.g. (1 variant -1)/(4 tokens -1)/*100% = 0%) and 100% variability when a different variation is produced for each token (e.g. (4 variants-1)/(4 tokens-1)% = 100%).
2. Like total token variability, this ratio was modified by subtracting 1 from the numerator and the denominator because at least 1 incorrect variant must occur per target that includes incorrect tokens. Variability among error productions only occurs when more than one incorrect variant is produced.

References

ALPERN, G., BOLL, T. and SHEARER, M., 1986, *Developmental Profile II* (Los Angeles, CA: Western Psychological Services).

Aram, D. M. and Glasson, C., 1979, Developmental apraxia of speech. Paper presented at the American Speech and Hearing Association Convention, Atlanta, GA.

Blakeley, R. W., 1980, *Screening Test for Developmental Apraxia of Speech* (Tigard, OR: C.C. Publications).

Carrow-Woolfolk, E., 1985, *Test for Auditory Comprehension of Language- Revised* (Allen, TX: DLM Teaching Resources).

Crary, M. A., 1993, *Developmental Motor Speech Disorders* (San Diego, CA: Singular Publishing Group).

Davis, B. L., Jakielski, K. J. and Marquardt, T. P., 1998, Developmental apraxia of speech: determiners of differential diagnosis. *Clinical Linguistics and Phonetics*, **12**, 25–45.

Dunn, L. M. and Dunn, D. L., 1981, *Peabody Picture Vocabulary Test- Revised* (Circle Pines, MN: American Guidance Service).

Ferry, P. C., Hall, S. M. and Hicks, J. L., 1975, 'Dilapidated' speech: developmental verbal dyspraxia. *Developmental Medicine and Child Neurology*, **17**, 749–756.

Goffman, L. and Smith, A., 1999, Development and phonetic differentiation of speech movement patterns. *Journal of Experimental Psychology: Human Perception and Performance*, **25**, 649–660.

Goldman, R. and Fristoe, M., 1986, *Goldman-Fristoe Test of Articulation* (Circle Pines, MN: American Guidance Service, Inc.).

Gordon-Brannan, M. and Hodson, B. W., 2000, Intelligibility/severity measurements of prekindergarten children's speech. *American Journal of Speech-Language Pathology*, **9**, 141–150.

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

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

Guyette, T. W. and Diedrich, W. M., 1981, A critical review of developmental apraxia of speech. In N. J. Lass (Ed.) *Speech and language: Advances in basic research and practice* (New York: Academic Press), pp. 1–49.

Hall, P. K., Jordan, L. S. and Robin, D. A., 1993, *Developmental Apraxia of Speech: Theory and clinical practice* (Austin, TX: Pro-Ed).

Hedrick, D. L., Prather, E. M. and Tobin, A. R., 1984, *Sequenced Inventory of Communication Development-Revised* (Seattle, WA: University of Washington Press).

Hodson, B. W., 1986, *Assessment of Phonological Processes-Revised* (Austin, TX: Pro-Ed).

Horwitz, S. J., 1984, Neurological findings in developmental verbal apraxia. *Seminars in Speech and Language*, **5**, 111–118.

Khan, L. M. L. and Lewis, N., 1986, *Khan-Lewis Phonological Assessment* (Circle Pines, MN: American Guidance Service, Inc.).

King, J. L., Jakielski, K. J. and Malone, K. A., 2001, Quantifying variability in developmental apraxia of speech: case study results. American Speech-Language-Hearing Association Convention, New Orleans, LA.

Marquardt, T. P., Sussman, H. M. and Davis, B. L., 2001, Developmental apraxia of speech: Advances in theory and practice. In D. Vogel and M. Cannito (Eds) *Treating Disordered Motor Speech Control*, second edition (Austin, TX: Pro-Ed).

Miller, N., 1992, Variability in speech dyspraxia. *Clinical Linguistics and Phonetics*, **6**, 77–85.

Morley, M. E., 1957, *The Development and Disorders of Speech in Childhood* (Baltimore, MD: Williams & Wilkins).

Murdoch, B. E., Porter, S., Younger, R. and Ozanne, A., 1984, Behaviours identified by South Australian clinicians as differentially diagnostic of developmental dyspraxia. *Australian Journal of Human Communication Disorders*, **12**, 55–70.

Newcomer, P. L. and Hammill, D. D., 1988, *Test of Language Development-Primary: Second Edition* (Austin, TX: Pro-Ed).

Robin, D. A., 1992, Developmental apraxia of speech: Just another motor problem. *American Journal of Speech-Language Pathology*, **1**(3), 19–22.

ROSENBEK, J. C. and WERTZ, R. T., 1972, A review of 50 cases of developmental apraxia of speech. *Language, Speech, and Hearing Services in Schools*, **5**, 13–22.

SCHUMACHER, J. G., MCNEIL, M. R., VETTER, D. K. and YODER, D. E., 1986, Articulatory consistency and variability in apraxic and non-apraxic children. Paper presented at the American Speech and Hearing Association Convention, Detroit, MI.

SHARKEY, S. G. and FOLKINS, J. W., 1985, Variability of lip and jaw movements in children and adults: implications for the development of speech motor control. *Journal of Speech and Hearing Research*, **28**, 8–15.

SHRIBERG, L. D., ARAM, D. M. and KWIATKOWSKI, J., 1997, Developmental apraxia of speech: II. Toward a diagnostic marker. *Journal of Speech Language and Hearing Research*, **40**, 286–312.

SKINDER, A., STRAND, E. A. and MIGNEREY, M., 1999, Perceptual and acoustic analysis of lexical and sentential stress in children with developmental apraxia of speech. *Journal of Medical Speech-Language Pathology*, **7**, 133–144.

SMITH, B. L., 1994, Effects of experimental manipulations and intrinsic contrasts on relationships between duration and temporal variability in children's and adults' speech. *Journal of Phonetics*, **22**, 155–175.

SMITH, B., MARQUARDT, T. P., CANNITO, M. P. and DAVIS, B. L., 1994, Vowel variability in developmental apraxia of speech. In J. A. Till, K. M. Yorkston and D. R. Beukelman (Eds) *Motor Speech Disorders: Advances in assessment and treatment*, (Baltimore, MD: Paul H. Brookes Publishing Co.), pp. 81–89.

SQUARE, P. A., 1994, Treatment approaches for developmental apraxia of speech. *Clinics in Communication Disorders*, **4**, 151–161.

STACKHOUSE, J. and SNOWLING, M., 1992, Developmental verbal dyspraxia: II. A developmental perspective on two case studies. *European Journal of Disorders of Communication*, **27**, 35–54.

SUSSMAN, H. M., DUDER, C., DALSTON, E. and CACCIATORE, A., 1999, An acoustic analysis of the development of CV coarticulation: a case study. *Journal of Speech Language and Hearing Research*, **42**, 1080–1096.

SUSSMAN, H. M., MARQUARDT, T. P. and DOYLE, J., 2000, An acoustic analysis of phonemic integrity and contrastiveness in developmental apraxia of speech. *Journal of Medical Speech-Language Pathology*, **8**, 301–313.

WILLIAMS, R., INGHAM, R. J. and ROSENTHAL, J., 1981, A further analysis for developmental apraxia of speech in children with defective articulation. *Journal of Speech and Hearing Research*, **24**, 496–505.

ZIMMERMAN, I. L., STEINER, V. G. and POND, R. E., 1979, *Preschool Language Scale-Revised Edition* (San Antonio, TX: The Psychological Corporation).