Speech Timing Variables in Children with Typical Speech Acquisition, Speech Delay, and Suspected Apraxia of Speech

[a]Jordan R. Green, [a]Lawrence D. Shriberg, & [b]Thomas F. Campbell

[a]University of Wisconsin-Madison [b]University of Pittsburgh

Purpose

•The research context for this report is the ongoing effort to identify one or more diagnostic markers for the putative child speech-sound disorder termed apraxia of speech. Classic descriptions of suspected childhood apraxia of speech include the construct of a "staccato-like" rhythmic quality, but few acoustic studies have attempted to quantify this percept (cf. Shriberg, Aram, & Kwiatkowski, 1997). Related variables in the adult neurogenic literature include the constructs of isochrony, scanning speech, syllable segregation, and more generally, abnormal speech timing (e.g., Ackerman & Hertrich 1994; Kent, Weismer, Kent, Vorperian, & Duffy, 1999). This regularity in speech timing may impact listeners' prosodic expectations and affect speech intelligibility.

•The purposes of the present report are (a) to describe a procedure to assess speech timing variables in conversational speech, (b) to provide speech timing reference data for 3- to 6-year-old speakers with typical speech and speech delay, and (c) to provide preliminary speech timing information on a group of children with suspected apraxia of speech. The reference data will be used to examine if children with suspected apraxia of speech exhibit a trend toward speech-timing regularity not observed in children with typical speech acquisition or speech delay.

Methods

Participants

• A total of 72 conversational speech samples were selected from the audiocassette archives of the University of Wisconsin-Madison Phonology Project. The ages of all speakers ranged from 3 to 6 years (3-4 years, 4-5 years, and 5-6 years).

Speakers with Normal Speech Acquisition (NSA)

• A total of 30 speakers, 5 boys and 5 girls at each of the three ages, scored above 95% on a severity metric termed the Percentage of Consonants Correct (PCCR) (Shriberg, Austin, Lewis, McSweeny, & Wilson (1997a) and were classified as Normal Speech Acquisition using the Speech Disorders Classification System (SDCS) module of the PEPPER suite (Shriberg, Allen, McSweeny, & Wilson, 2000).

Speakers with Speech Delay (SD)

• A total of 30 children, 5 boys and 5 girls at each of the three ages, scored below 85% on the PCCR metric and were classified as having Speech Delay (SD) using the SDCS metric cited above.

Speakers with Suspected Apraxia of Speech (sAOS)

• A total of 12 children with Suspected Apraxia of Speech (sAOS) were selected from approximately 100 children with clinical referral histories consistent with sAOS. Their lexical and sentential stress characteristics were considered inappropriate by consensus transcription of two experienced researchers using auditory-perceptual criteria (McSweeny & Shriberg, 2001).

Speech Processing

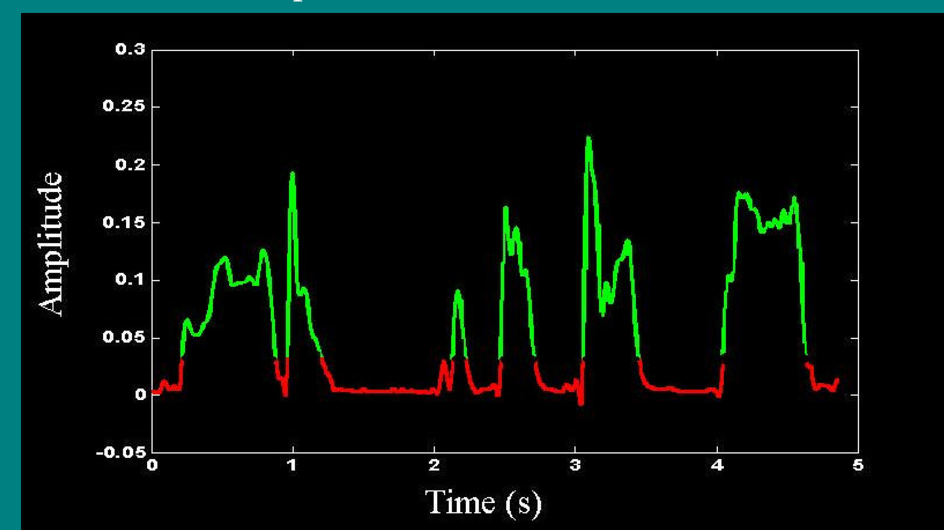
Digitizing and Segmentation

• The audiocassette archive included conversational speech samples from each of the 72 speakers. A subset of 24 utterances from each sample was selected using criteria for prosody-voice coding (McSweeny & Shriberg, 2001). Each child's utterances were digitized at 44.1 KHz/s (16 bit).

• Computer-assisted detection methods were developed to determine the onset and offset of pause and speech events within each of the 24 utterances in each conversational speech sample. For this procedure, the investigators were required to identify a region on a rectified and digitally filtered display of each waveform that contained the largest amplitude "pause" event. The maximum value of the selected pause region was used to establish a threshold for separating "speech" from "pause" events.

• Boundaries associated with each pause event were identified as values in the rectified waveform below the threshold, and speech events were identified as values that were above the threshold (see Figure 1). This method captures the general energy distribution within each utterance by identifying high-energy portions in the rectified waveform as speech events and low energy portions as pause events. The time coordinate values associated with each speech and pause event were exported to a database.

• Figure 1: A computer-assisted method was used to automatically identify "pause" and "speech" events in each utterance. For this utterance, red sections of the rectified waveform were identified as *pause* events and green sections were identified as *speech* events.



• Pause and speech durations less than 100 ms were excluded from the data set to reduce the effects of short pause events related to stop closure and short speech events related to burst release on the present analysis. Thus, the pauses analyzed in this report were located both within and between words, and at prosodic phrases and syntactic boundaries (i.e., phrase, clause, and sentence).

• Preliminary examination of the data indicated that length of utterance varied considerably across speakers, including speakers in the same age groups. A program in the PEPPER suite was used to divide speakers into two groups according to their Average Words Per Utterance (AWU).

• **Table 1** provides a summary of the numbers of speech events and pause events for speakers in each of the three groups as a function of AWU (1.5 – 3.9 ; 4.0 – 8.0)

AWU	Group	Pause Events	Speech Events
1.5 - 3.99	NSA (n=10)	481	796
	SD (n=15)	572	1097
	sAOS (n=6)	563	330
4-0 - 8.0	NSA (n=20)	1278	1749
	SD (n=15)	919	1221
	sAOS (n=6)	544	477

Analyses

• Analyses were designed to (1) characterize the distributional characteristics of pause and speech events, and (2) detect the variability of pause and speech events within each utterance. Pause and speech durations were pooled across subjects and cross-tabulated by AWU and Speaker Group.

• **Analysis 1:** Summary statistics were computed for each subject's distributions of pause and speech durations.

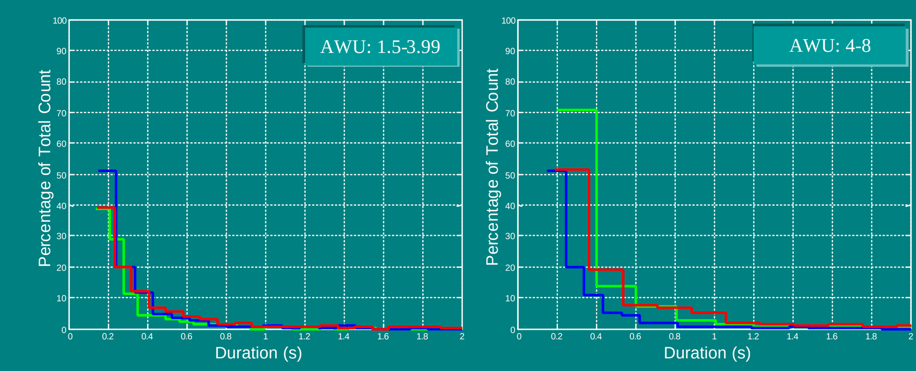
• **Analysis 2:** A measure of *within-utterance* variability in pause and speech durations was also derived. The coefficient of variation (CV = SD/Mean) was computed across all of the durations within an utterance for each event type (i.e., speech and pause). This transformation adjusted for changes in event durations that were related to variations in speech rate across utterances and speakers. CV values were averaged across utterances for each subject. Panel A contains an example of the CV calculation for a single utterance.

Results

Analysis 1. Summary statistics. Tables 2a and 2b, and Figures 2a and 2b are displays of the distributions of unadjusted pause and speech durations for speakers in the three speaker groups.

Table 2a & Figure 2a. Pause Data

AWU	Group	IQR	St Dev	Mean	Median	Skew	Kurtosis
1.5-3.99	NSA	0.15	0.23	0.27	0.19	3.84	24.52
	SD	0.18	0.26	0.28	0.19	3.80	25.62
	sAOS	0.39	0.50	0.45	0.27	3.51	23.83
4.0-8.0	NSA	0.20	0.41	0.33	0.20	5.62	53.76
	SD	0.18	0.26	0.28	0.19	3.80	25.62
	sAOS	0.39	0.50	0.45	0.27	3.51	23.83



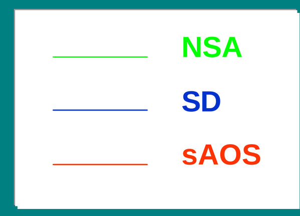
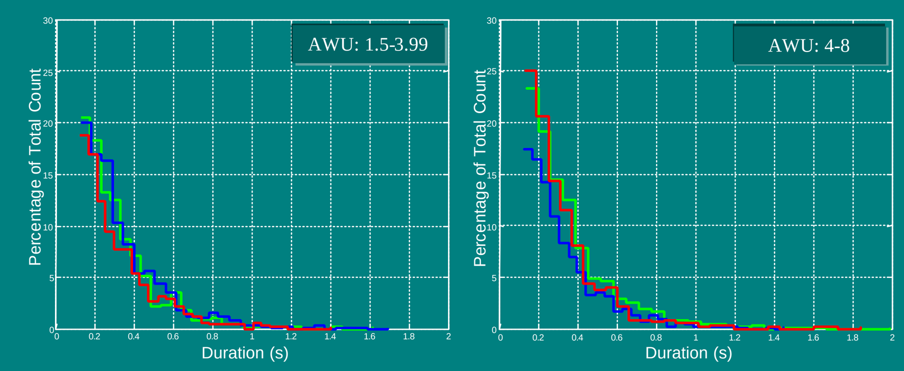


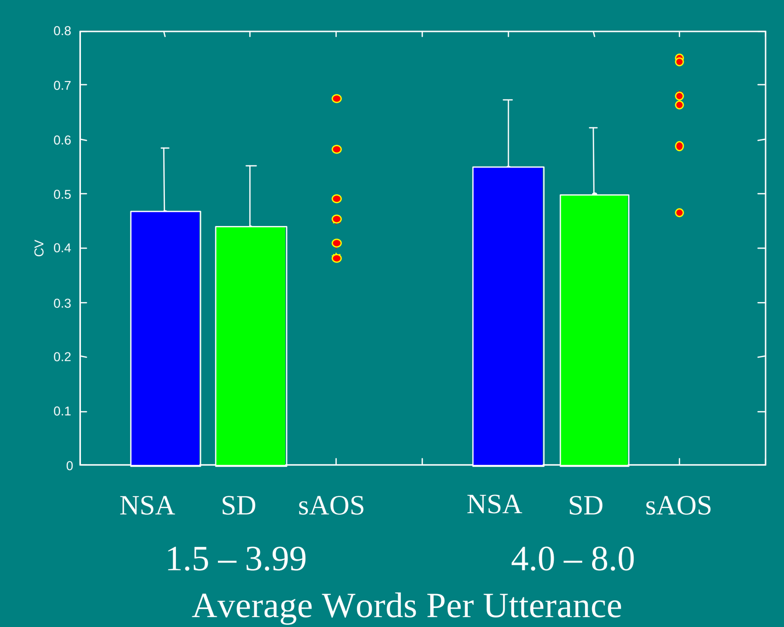
Table 2b and Figure 2b. Speech Data

AWU	Group	IQR	St Dev	Mean	Median	Skew	Kurtosis
1.5-3.99	NSA	0.21	0.19	0.30	0.25	2.04	9.50
	SD	0.23	0.21	0.32	0.25	2.06	9.08
	sAOS	0.22	0.19	0.30	0.24	1.74	6.84
4.0-8.0	NSA	0.23	0.22	0.32	0.26	2.04	8.90
	SD	0.22	0.20	0.31	0.24	1.88	7.59
	sAOS	0.20	0.22	0.30	0.23	2.75	14.55



Analysis 2. Within utterance variability in pause- and speech-event durations. Figures 3a and 3b are displays of the average within-utterance CVs for pause- and speech-event durations for speakers within each speaker group.

Figure 3b. Coefficient of Variation for Pause Data

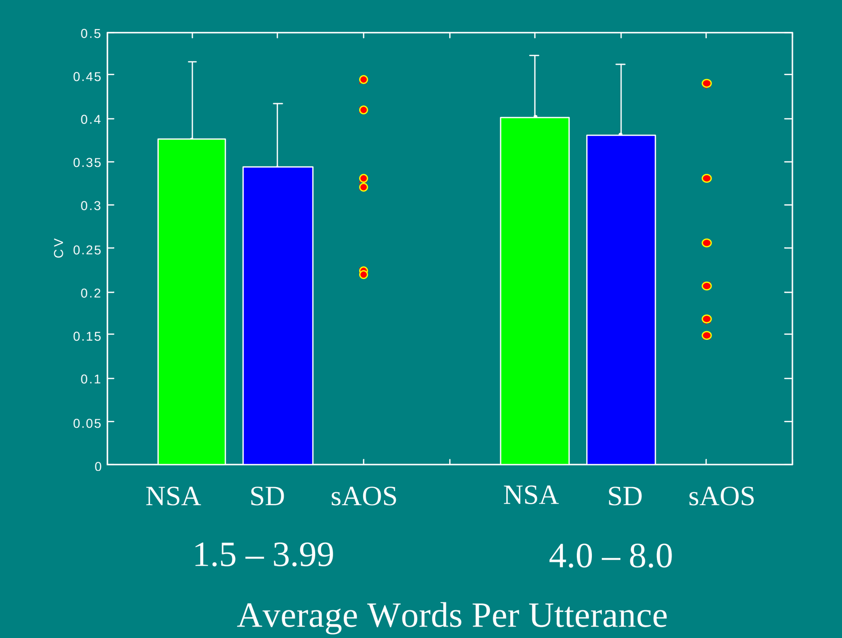


Panel A. Example of Data Reduction

•Consider an utterance with 5 speech events that has the following durations:
0.55, 0.32, 0.35, 0.22, 0.17 (s)

	Durations
	0.55
	0.32
	0.35
	0.22
	0.17
SD	0.09
Mean	0.20
CV (SD/Mean)	0.46

Figure 3b. Coefficient of Variation for Speech Data



Primary Findings

- Pause-event durations within an utterance are significantly more variable in children with sAOS than children with NSA and SD (Tukey Test, $p < .05$, for all comparisons). There were no statistically significant differences in pausal variability between NSA and SD speakers.
- For the 4-8 AWU group only, speech-event durations within an utterance were significantly less variable in children with sAOS than children with NSA and SD (Tukey Test, $p < .05$). There were no statistically significant differences in speech-event variability between NSA and SD speakers.

Discussion

Method

The method developed for quantifying speech signal variability appeared to capture temporal differences in conversational speech among these three speaker groups.

Diagnostic Marker

The finding of temporal variability in pause-event durations in speakers with sAOS, and the trend for temporal regularity in their speech-event durations, adds support for the possibility of a prosodic-based diagnostic marker for these children. Only the trend for decreased variability of speech-event durations within an utterance is consistent with perceptual descriptions of excessive/equal stress (Shriberg et al., 1997). Excessive/equal stress may be one perceptual correlate of isochrony, scanning speech, or syllable segregation. The increased variability in pause-event durations within utterances is consistent with observations of long and variable intersyllabic pauses observed in adults with AOS, but is inconsistent with a strict definition of isochrony.

Speculation

The divergent trends for pause- and speech-event durations in the sAOS group suggest that these events are represented independently during phonetic encoding. The increased variability in pause events in children with sAOS might be the consequence of a phonetic-encoding deficit.

In contrast, the tendency toward regularity in speech events might be either a behavioral manifestation of a reduced capacity for phonetic encoding, or a simplification strategy used by the child to make speech timing more predictable.

References

Ackermann H., & Hertrich I. (1994). Speech rate and rhythm in cerebellar dysarthria: An acoustic analysis of syllabic timing. *Folia Phoniatrica et Logopedica*, *46*, 70-78.

Kent R. D., Weismer G., Kent, J. F., Vorperian, H. K., & Duffy J. R. (1999). Acoustic studies of dysarthric speech: Methods, progress, and potential. *Journal of Communication Disorders*, *32*, 141-80.

McSweeny, J. L., & Shriberg, L. D. (2001). Clinical research with the prosody-voice screening profile. *Clinical Linguistics and Phonetics*, *15*, 505-528.

Shriberg, L. D., Allen, C. T., McSweeny, J. L., & Wilson, D. L. (2001). PEPPER: Programs to examine phonetic and phonologic evaluation records [Computer software]. Madison, WI: Waisman Center, University of Wisconsin.

Shriberg, L. D., Aram, D. M., & Kwiatkowski, J. (1997). Developmental apraxia of speech: III. A subtype marked by inappropriate stress. *Journal of Speech, Language, and Hearing Research*, *40*, 313–337.

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

Acknowledgements

• Thanks to Connie Nadler, Jane McSweeny, Alison Scheer, and Katherina Hauner for their important contributions to this study. This research was supported by NIH grants DC0046 and DC04643. Copies of this poster can be downloaded at www.waisman.wisc.edu/phonology.

• This poster was presented at the Conference on Motor Speech (March 2002), Williamsburg, VA.