Published in final edited form as:

Clin Linguist Phon. 2019 ; 33(5): 479–495. doi:10.1080/02699206.2018.1550813.

Word-level Prosodic Measures and the Differential Diagnosis of Apraxia of Speech

Katarina L. Haley and Adam Jacks

Department of Allied Health Sciences, University of North Carolina at Chapel Hill School of Medicine, USA

Abstract

Impaired production of prosody is considered a primary diagnostic criterion for apraxia of speech. In this study, we examined diagnostic relevance for five word-level prosody measures. Seven speakers with AOS, nine with aphasia and no AOS, and 19 age-matched neurotypical controls produced nine words consisting of three or four syllables. Lexical stress indices were computed based on relative values for duration, fundamental frequency, and intensity across pairs of unstressed-stressed syllables with varying intrinsic vowel duration contrast patterns. A simple average syllable duration measure was also obtained. AOS speakers differed from the other two groups on three metrics that were solely or primarily derived from duration measures. The degree of diagnostic overlap was smallest for the syllable duration metric, which also showed the strongest inter-observer reliability and most complete overlap between neurotypical speakers and speakers with aphasia and no AOS. Vowel intrinsic durational properties affected lexical stress metrics significantly, indicating a need to select word targets purposefully. Based on these results, it appears that the average syllable duration metric is a more stable and informative alternative for differential diagnostic purposes. The results will, however, need to be replicated in a larger sample.

Keywords

Apraxia of speech; aphasia; pairwise variability index; word syllable duration; word prosody

Acquired apraxia of speech (AOS) is defined as a motor speech disorder that results from left hemisphere lesions and affects the planning or programming of speech movements. It is diagnosed based on a collection of features that are evident during most speech production tasks and differentiates the disorder, to varying degrees, from aphasia without AOS. Slow and halting articulation is one of the most robust AOS features and impaired prosody is therefore considered a primary diagnostic criterion. The feature has held this status for over 50 years (Ballard et al., 2015; Darley, 1968; McNeil, Robin, & Schmidt, 2009; Wertz, LaPointe, & Rosenbek, 1984). Specific manifestations include sound prolongation, syllable and sound segmentation, and stress contrast reduction (Haley, 2002; Kent & Rosenbek,

Please address correspondence to: Katarina Haley, 3124 Bondurant Hall, University of North Carolina at Chapel Hill, Campus Box 7190, Chapel Hill, NC 27599-7190 USA, Katarina_Haley@med.unc.edu, Telephone: 919-445-0341, Fax: 919-966-0100.

Statement of interest: The authors report no conflicts of interest.

1982, 1983; Odell, McNeil, Rosenbek, & Hunter, 1990; Odell, McNeil, Rosenbek, & Hunter, 1991; Seddoh et al., 1996; Skenes, 1987).

When pronounced, the prosodic signs of AOS are usually unmistakable to trained diagnosticians, though variations in severity may elude detection. When they are moderate or mild, perceptual evaluation becomes far more challenging and it may be unclear whether the variation is normal or impaired. Less experienced clinicians may not fully appreciate characteristic prosodic features and fail to recognize them even when they are pronounced. For these reasons, there is a need to supplement perceptual impression with quantitative information that is capable of informing diagnostic differentiation. In the area of prosody, acoustic measures are ideally suited and relatively straightforward to derive. Unfortunately, there has so far been limited effort to evaluate the relative advantages of candidate metrics relative to different assessment purposes and to organize measurement procedures for clinical application. In the present study, we evaluate the degree to which five measures of lexical prosody differentiate AOS from normal speech production and from speech production in aphasia that is not complicated by AOS. We selected these five measures based on previously demonstrated promise and suitability for analysis at the single word level.

Though not reflective of communication in daily life, single word repetition has important advantages for our purposes relative to sentence or discourse production, structured word retrieval, and a host of other elicitation methods. Among these advantages are constrained linguistic influence, standardized production targets, and adequate production challenge. Most people with acquired AOS have coexisting aphasia and this aphasia is often characterized by varying degrees of inefficiency in syntax production. Inefficient production of phrases and sentences have secondary effects on prosody and, therefore, confounds analyses that target the prosodic consequences of motor planning or programming impairment. Additionally, hesitancy related to inefficient lexical retrieval or orthographic decoding is circumvented in a word repetition task. Finally, because production content can be controlled, it is possible to prevent strategic avoidance of complex targets as well as the experience of repeated failure.

Ratios for lexical stress

Prosodic variation extends suprasegmentally across syllable sequences of varying length. For most purposes, evaluation requires a sample of at least two syllables and it is often desirable to target words with three or more syllables. Lexical stress, for example, is expressed by relatively longer duration, greater intensity, and higher fundamental frequency in stressed syllables relative to unstressed syllables. If the lexical stress contrast is reduced in people with AOS, then the acoustic difference between stressed and unstressed syllables should be lower than normal. The pairwise variability index (PVI) was originally conceptualized as the normalized duration difference between adjacent syllables in connected speech and used to characterize speech rhythm and stress in neurotypical speakers who use different languages and dialects (Low, Grabe, & Nolan, 2000). Direct application to running speech has been possible for communication disorders that do not primarily affect linguistic fluency (Knight & Cocks, 2007; Maruthy, Venugopal, & Parakh, 2017) and, when

modified for comparisons at the word level, the PVI has contributed to research addressing lexical stress development in children (Arciuli & Ballard, 2017; Arciuli & Colombo, 2016; Ballard, Djaja, Arciuli, James, & van Doorn, 2012) and the differentiation between AOS and aphasia (APH).

Most attention has been given to the PVI for durational values (PVI_{dur}), with only a few studies extending the application to intensity and fundamental frequency variations. Relative to the differentiation between AOS and APH, comparisons between the duration of unstressed and stressed syllable nuclei have been most fruitful for words with a weak-strong stress pattern, such as the first and second syllables in words like *catastrophe*, *potato*, and *detergent* (unstressed first syllable, stressed second syllable). The stressed syllable is normally longer than the unstressed syllable and the greater the stress difference, the higher the PVI_{dur} values. Consistent with a pattern of reduced stress contrasts, lower PVI_{dur} values have been noted for speakers with AOS than for speakers with APH and neurotypical controls (Ballard et al., 2014, 2016; Duffy et al., 2015; Vergis et al., 2014; Whitwell et al., 2017). Vergis and colleagues observed that the difference may be more evident when the target word is embedded in a sentence context than when it is produced in isolation and that the effect is less clear in target words with the trochaic strong-weak stress pattern that is most typical in English (e.g. the first two syllables of the word ‘thickening’). These observations indicate that PVI_{dur}, like durational relationships more broadly, is affected by factors beyond lexical stress production and likely involves complexity issues pertinent to speech planning or programming.

Other acoustic dimensions of lexical stress have received more limited attention and appear less promising. Even though speech output in AOS has been observed to manifest with reduced intensity variation from syllable to syllable (Kent & Rosenbek, 1983) and PVI for peak intensity has differentiated lexical stress in children and adults (Ballard, Djaja, Arciuli, James, & van Doorn, 2012), it has so far not differentiated speakers with acquired AOS and aphasia from those with aphasia only or from neurotypical adults (Ballard et al., 2014; Vergis et al., 2014). Similarly, when derived from fundamental frequency data, PVI has yielded equivocal results in terms of sensitivity to lexical stress (Ballard et al., 2012; Ballard, Robin, McCabe, & McDonald, 2010). Suitability is, therefore, questionable for applying a fundamental frequency PVI metric to the AOS-APH differentiation.

Even if further studies indicate that the diagnostic value of lexical stress correlates for intensity and fundamental frequency by themselves have limited effect in differentiating AOS from aphasia without AOS, a composite metric that integrates all three dimensions may enhance precision, given that trading relations are often observed among acoustic stress correlates and have been observed in speakers with acquired AOS (Marquardt, Duffy, & Cannito, 1995). One such weighted composite metric is the lexical stress ratio (LSR; Shriberg et al., 2003; van Santen, Prud’hommeaux, & Black, 2009). The LSR was developed as a potential acoustic marker of lexical stress for speakers with childhood apraxia of speech (CAS), which has been characterized by abnormal stress at both word and phrase levels (Shriberg, Aram, & Kwiatowski, 1997a, 1997b, Odell & Shriberg, 2001). The measure is a weighted linear combination of ratios of acoustic metrics from stressed and unstressed syllables, specifically including duration, fundamental frequency, and amplitude. The

weightings and factors were obtained from a factor analysis of trochee words (strong-weak stress pattern) produced by 11 children with suspected CAS and 24 children with speech delay. Although the measure was interpreted as being sensitive to stress differences in children with CAS, the relationship between stress and LSR measurements in Shriberg's sample is unclear, and there is not a known normative range for this measure. Notwithstanding these limitations, to our knowledge LSR is the only measure that combines the three primary acoustic correlates of stress and, as such, it may be suitable for indexing lexical stress reduction in acquired AOS and potentially differentiate it from aphasia without AOS.

Duration of segments and pauses

Overall word duration obviously increases when word length increases in number of syllables. In neurotypical speakers, this increase is moderated by a concurrent duration reduction for constituent syllables (Klatt, 1976; Lehiste, 1972). Speakers with AOS maintain this reduction (Baum, 1992; Collins, Rosenbek, & Wertz, 1983), but with a smaller magnitude than neurotypical controls and speakers with aphasia and no AOS, such that overall duration of words with three or more syllables is significantly longer for speakers who have AOS than for those who have aphasia and no AOS (Collins et al., 1983; Haley & Overton, 2001).

AOS is characterized by speech sound prolongation as well as intersegmental pauses (Seddoh et al., 1996; Skenes, 1987; Strand & McNeil, 1996). Both features contribute to a metric, known as the word syllable duration (WSD), which is calculated as word duration divided by number of syllables produced (Haley, Jacks, de Riesthal, Abou-Khalil, & Roth, 2012). The metric is essentially the inverse of unadjusted articulation rate in multisyllabic words. Based primarily on face validity, Haley and colleagues have used WSD as a quantitative proxy criterion for prosodic impairment in AOS (Cunningham, Haley, & Jacks, 2016; Haley, Jacks, & Cunningham, 2013; Haley, Jacks, Richardson, & Wambaugh, 2017). It remains unknown how the measure performs as a dependent rather than grouping variable. Of note, the distribution pattern appears to be continuous in stroke survivors and has no clear discontinuity that would unequivocally separate impaired from normal performance (Haley et al., 2017).

Intrinsic duration of vowels

There is inherent variability in speech production. Occasionally, neurotypical speakers also prolong speech sounds and syllables and insert pauses within words. However, in comparison to AOS speakers, such occurrences are relatively rare. To detect the persistent patterns that are indicative of impairment, it is necessary to sample several word productions and integrate measurements across them. When eliciting a speech sample for analysis, it is advantageous to consider how related qualities of the target words may influence prosodic measures. This is particularly important when—as is sometimes the case—only a small number, or even a single word is available for sampling. We remarked previously that PVI_{dur} for a weak-strong stress pattern differentiates better between AOS and aphasia without AOS than a strong-weak stress pattern (Ballard et al., 2014; Vergis et al., 2014) and

that WSD differences are enhanced by 3- and 4- syllable words relative to 2-syllable words (Collins et al., 1983; Haley & Overton, 2001). In this study, we examine the effect of intrinsic vowel duration on lexical prosodic measures.

Vowel length is moderated by several factors. Two of those factors are speaking rate and lexical stress, both of which involve lexical prosody and are of interest to our ambition of quantifying word-level prosody. Another factor is intrinsic vowel duration. Consider, for example, the word ‘catastrophe’, for which the first two syllables have previously been used to index lexical stress (e.g. Duffy et al., 2017). The target word is a good choice in that it has a weak-strong stress pattern and syllable nuclei with relatively clear acoustic boundaries relative to surrounding obstruent consonants. However, the duration contrast is not limited to lexical stress and, therefore, the PVIDur signifies more than lexical stress in this word. The first vowel in ‘catastrophe’ is a neutral vowel with short intrinsic duration in American English, whereas the second vowel /æ/ has an intrinsically long duration. Across a group of speakers, the average difference between these vowel qualities approximates 100 ms when produced in a monosyllabic /h_d/ context (Hillenbrand, Getty, Clark, & Wheeler, 1995). If vowel qualities were maintained, but our target word was produced with equalized stress, the PVIDur would still have some magnitude due to the intrinsic vowel difference. When lexical stress contrast (unstressed-stressed) is added to the intrinsic vowel duration contrast (short-long), the PVIDur magnitude would operate congruently and the combined factors would potentially magnify the duration contrast.

For other words, the duration contrast may operate in opposite directions and potentially neutralize the stress contrast and reduce PVIDur magnitude. For example, in the word ‘perception’ the rhotic vowel for the first syllable is unstressed but has an intrinsically long duration, whereas the lax vowel /e/ in the second syllable is stressed but intrinsically short (Hillenbrand et al., 1995). Whereas there is no perceptual penalty of prolonging an intrinsically long vowel such as /æ/, there are inherent limitations in the extent to which prolongation of the intrinsically short vowel /e/ is possible without compromising vowel quality (Klatt, 1976). For these reasons, we would expect PVIDur in neurotypical speakers to be of greater magnitude in words like ‘catastrophe’ where intrinsic vowel duration and lexical stress contrast are congruent than in words like ‘perception’ where they are opposite.

To evaluate the relative merits of different word-level prosodic measures, we asked two questions: 1) Which metric differentiate AOS most completely from aphasia without AOS and from normal speech? 2) Does intrinsic vowel duration systematically affect ratio metrics that have been used to index lexical stress?

Methods

Participants

Participants were 16 people with aphasia and 19 age-matched neurotypical control (NTC) participants without known speech difficulties. They were native speakers of English and reported no previous history of speech, language or neurological impairment. All participants passed a hearing screening at 40 dB HL in the better ear for 1000 Hz and 2000 Hz (Ventry & Weinstein, 1983). The study was approved by the university’s Institutional

Review Board and all participants provided signed informed consent. Seven of the 35 identified as African American, one reported Hispanic/Latino ethnicity, and the remaining participants reported white non-Hispanic/Latino background. They were at least 4 weeks post stroke onset, had the ability to repeat single words, and did not carry a diagnosis of progressive neurologic disorder. A motor speech evaluation was audio-recorded and reviewed by the authors for the purpose of differential diagnosis between aphasia with AOS and aphasia without AOS and verification that participants had no or potentially very mild dysarthria of the unilateral upper motor neuron type (Duffy, 2013). These speech samples were reviewed by the two authors, who are certified and licensed speech-language pathologists with extensive training in the assessment and diagnosis of acquired neurogenic communication disorders. Their diagnostic agreement was 100% concerning AOS diagnosis and absence of significant dysarthria. Independently of this clinical impression, the audio-recordings were quantified by a team of research assistants, using a series of perceptual and acoustic metrics, listed in table 1.

Seven of the participants with aphasia were diagnosed with AOS. Based on the Aphasia Diagnostic Profiles (ADP; Helm-Estabrooks, 1992), four profiled with Broca's aphasia and three profiled with borderline fluent aphasia. Single word intelligibility for these speakers ranged from 51% to 77%, indicating moderate sound production severity. Nine participants were diagnosed with aphasia without AOS (APH). Of these, four had an ADP profile of conduction aphasia, one had a profile of anomic aphasia, three had a profile of borderline fluent aphasia, and one presented with aphasia that was too mild to be classified. Speech intelligibility scores in the APH group scores ranged from 77% to 94%, indicating moderate to mild sound production impairment. Frequency of error type was derived from narrow phonetic transcriptions of the motor speech evaluation. The phonemic edit distance ratio—an index of substitution, addition, and omission frequency (Smith, Haley, & Wambaugh, 2017) and distortion frequency was higher for the AOS speakers than for the APH speakers. A clinical measure of WSD was calculated from six multisyllabic words (gingerbread, television, thickening, zippering, jabbering, flattering) by measuring word duration in milliseconds and dividing this duration by the number of syllables produced. This mean WSD metric was greater and more variable in the AOS group ($M=428.2$ ms, $SD=160.1$ ms) than in the APH ($M=236.6$ ms, $SD=36.6$ ms) and NTC ($M=232.0$ ms, $SD=31.1$ ms) groups. As a clinical measure of PVIDur, we measured the duration of the first and second syllable nuclei in the word 'catastrophe.' On the motor speech evaluation protocol, this word was repeated five times consecutively, so we expressed PVIDur as the mean across each participant's repeated attempts. PVIDur was smallest in the AOS group ($M=11.6$, $SD=60.1$), intermediate in the APH group ($M=30.2$, $SD=28.2$), and largest in the NTC group ($M=113.0$, $SD=15.5$). Four speakers (three in the AOS group and one in the APH group) were unable or unwilling to repeat this word.

Speech samples and acoustic analysis

All 35 speakers produced three four-syllable words (conversation, combination, publication) and six three-syllable words (collection, condition, connection, position, perception, permission). Each target word included a sequence of an unstressed syllable followed by a stressed syllable, but the intrinsic vowel duration varied. The variation is summarized in

table 2 by the **vowel duration contrast** between the first and second analysed syllable nucleus (congruent with lexical stress, neutral relative to lexical stress, opposite to lexical stress).

For the two words with congruent difference (combination, publication), the intrinsic vowel duration contrast was in the same direction as the lexical stress contrast (short/unstressed in the second word nucleus followed by long/stressed diphthongized nucleus in the third word nucleus). We predicted the duration contrast would have the greatest magnitude for these targets. For five words (conversation, collection, condition, connection, position), intrinsic vowel duration was equivalent for the analysed nucleus pair. With a neutral intrinsic vowel duration contrast, an observed duration difference would be more specific to lexical stress. Therefore, we predicted the duration difference magnitude would be intermediate. For the two remaining words (perception, permission), the intrinsic duration of the (rhotic) nucleus was in the opposite direction of the stress contrast (intrinsically long/unstressed rhotic nucleus followed by intrinsically short/stressed nucleus). We anticipated the duration contrast would be smallest for these targets.

Speech samples were recorded in a quiet room via a head-mounted microphone (AKG-C420). Pre-recorded auditory models were presented for repetition via external speakers. The volume was set at a level that was determined comfortable by each participant. Recorded audio signals were segmented using Praat software (Boersma & Weenink, 2017) by manually placing boundaries for the word and the syllable nuclei in TextGrids. Duration, median fundamental frequency, and median intensity (in dB) were obtained from the resulting segments. All target approximations were analyzed as long as they were produced with the same number of syllables as the target. As predicted by the sound segment error frequencies reported in table 1, the approximations included both phonemic and phonetic errors. Our rationale was that inclusion of words with sound errors was necessary to ensure adequate and representative sampling, since these errors are typical of the speech output.

Five word-level prosodic measures were derived from this segmentation. The measures of lexical stress included three variants of the pairwise variability index (PVI; Ballard et al., 2014, 2016; Duffy et al., 2017; Low et al., 2000; Vergis et al., 2014), one for each acoustic dimension. In addition, we computed the lexical stress ratio (LSR) by integrating the three acoustic dimensions (Shriberg et al., 2003). The final prosodic metric was word syllable duration (WSD; Cunningham et al., 2016; Haley et al., 2012; 2013; 2017).

Pairwise variability index (PVI).—The pairwise variability index (PVI) was calculated as the normalized difference between unstressed and stressed syllables for duration, fundamental frequency, and intensity, according to the following formula, where x_1 is the acoustic value for the unstressed syllable and x_2 is the acoustic value for the stressed syllable:

$$PVI = \frac{(x_2 - x_1)}{(x_1 + x_2)/2}$$

Lexical stress ratio (LSR).—The lexical stress ratio (LSR; Shriberg et al., 2003) is a composite measure of duration, amplitude, and fundamental frequency, with duration as the most prominent component. Calculation of LSR proceeded in the following steps (see also van Santen et al., 2009):

1. Amplitude area and pitch area were calculated for each syllable, based on the products of amplitude and pitch with the respective syllable durations, where D is duration, F_0 is fundamental frequency, and A is amplitude for syllables $i = 1$ (unstressed) and 2 (stressed)

$$\text{Amplitude area} = D_i A_i; \text{Pitch area} = D_i F_{0i}$$

2. A ratio was formed for amplitude area, pitch area, and duration, combining the areas computed in step 1:

$$\text{Amplitude area ratio} = \frac{D_1 A_1}{D_2 A_2}; \text{Pitch area} = \frac{D_1 F_{01}}{D_2 F_{02}}; \text{Duration ratio} = \frac{D_1}{D_2}$$

3. Ratios from step 2 were combined with weights α , β , and γ determined from Shriberg et al. as $\alpha = 0.507$, $\beta = 0.490$, and $\gamma = 0.303$

$$LSR = \alpha \frac{D_1 A_1}{D_2 A_2} + \beta \frac{D_1 F_{01}}{D_2 F_{02}} + \gamma \frac{D_1}{D_2}$$

Word syllable duration (WSD).—The mean word syllable duration was calculated across words, based on total word duration divided by the number of syllables produced (Haley et al., 2012).

Inter-rater reliability.—To estimate inter-rater reliability, a second coder segmented approximately 20% of the sample. Speakers from all three participant groups were included in proportionate numbers (4 NTC speakers, 2 speakers with APH, 2 speakers with AOS. All words were segmented for each of these speakers. We used single measure absolute agreement two-way mixed effects intraclass correlations (ICC) to estimate inter-rater reliability. Our first estimate was for the duration, intensity, and fundamental frequency of the vowel nuclei in the syllables of interest (see table 3). These were all 0.90 or higher, indicating excellent agreement (Koo & Li, 2016). Because precision of constituent measures does not guarantee precision of derived metrics (Hallgren, 2012), we also assessed agreement for each of the word-level prosodic measures, using the same ICC model. Agreement was excellent for WSD (ICC = 0.99), however it was only moderate for most of the derived lexical stress metrics (ICCs from 0.65 to 0.70; see table 3), with the exception of PVI for fundamental frequency (ICC = 0.96). We will return to this observation in the discussion of results.

Statistical analysis.—Acoustic measures for each word and participant were entered into 2-way ANOVAs, with speaker group (AOS, APH, NTC) and vowel duration contrast

(congruent, neutral, opposite relative to lexical stress) as factors. Effects were modeled for group and vowel duration contrast as well as the group by duration contrast interaction. For significant main effects without an interaction, pairwise t-tests were used to examine differences among levels. For measures with significant interactions, post-hoc analysis consisted of one-way “slice” ANOVAs to separately examine the effect of vowel duration contrast on each group. When interactions were not adequately explained by the one-way ANOVAs, pairwise t-tests were used to examine differences.

Results

Group differences.

Table 4 presents results from the five 2-way ANOVAs, one for each prosodic measure. PVI for fundamental frequency and intensity did not differentiate among the groups. Further analysis was therefore restricted to the three metrics that showed a significant group difference. These were PVI_{dur} ($F_{2, 277} = 73.73, p < .0001$), LSR ($F_{2, 277} = 85.80, p < .0001$), and WSD ($F_{2, 277} = 183.04, p < .0001$)—the three metrics that were solely or primarily based on duration measures.

Figure 1 shows the distribution of scores within each speaker group and intrinsic vowel duration contrast. Results are presented in separate panels for each duration-based prosodic measure. Note that the two lexical stress metrics in the top and middle panels distribute in opposite directions. Lower values for PVI_{dur} and higher values for LSR indicate reduced difference between the analyzed syllable pairs, as would be expected when lexical stress and intrinsic vowel duration contrast is diminished. Results for WSD are presented in the bottom panel. High WSD values indicate slow articulation rate as would be predicted in speakers with AOS.

Post hoc pairwise comparisons showed that group effects reflected differences for the AOS group relative to both the APH group (PVI_{dur}: $t_{283} = 8.09, p < .0001$; LSR: $t_{283} = 9.10, p < .0001$; WSD: $t_{283} = 17.34, p < .0001$) and the NTC group (PVI_{dur}: $t_{283} = 11.19, p < .0001$; LSR: $t_{283} = 12.03, p < .0001$; WSD $t_{283} = 19.46, p < .0001$). There was also significant differences between the APH and NTC groups for PVI_{dur} ($t_{283} = 2.56, p = .011$) and LSR ($t_{283} = 2.21, p = .03$). There was no difference between the APH and NTC groups for WSD ($t_{283} = .05, p = 0.96$).

Intrinsic vowel duration effects.

Omnibus and post hoc analyses confirmed that intrinsic vowel duration affected the two lexical stress metrics, but not WSD. As shown in table 4, the 2-way ANOVA for PVI_{dur} indicated an effect of vowel duration contrast ($F_{2, 277} = 26.12, p < .0001$) and vowel duration contrast by speaker group interaction ($F_{4, 277} = 3.62, p = .0068$). The interaction reflects a significant effect of vowel duration contrast for APH and NTC speakers ($F_{2, 277} = 14.97, p < .0001$; $F_{2, 277} = 40.07, p < .0001$, respectively) but not for the AOS speakers ($F_{2, 277} = 2.16, p = 0.12$).

Results for LSR were similar to those for PVI_{dur}, exhibiting more reliable effects of vowel contrast for NTC and APH speakers than in the AOS group. The 2-way ANOVA (table 4)

was significant for vowel duration contrast ($F_{2, 277} = 24.60, p < .0001$), and group by vowel duration contrast interaction ($F_{4, 277} = 3.13, p = .0153$). Since one-way ANOVAs showed an effect of vowel duration contrast for all three speaker groups (NTC, $F_{2, 277} = 18.47, p < .0001$; APH, $F_{2, 277} = 7.12, p = .001$; AOS, $F_{2, 277} = 11.24, p < .0001$), further pairwise comparisons were completed within each group to explore the interaction. These post hoc tests showed robust effects of vowel duration contrast on LSR for the NTC group (e.g. significant differences among all conditions, $p_s < .0001$), while only two comparisons differed for the APH group (congruent vs. neutral, $t_{73} = 3.78, p < .001$; neutral vs. opposite $t_{73} = 3.85, p < .001$), and one comparison differed for AOS group (neutral vs. opposite, $t_{43} = 2.40, p = .02$).

There was no effect of vowel duration contrast for the WSD metric. The 2-way ANOVA indicated statistical significance for group only. Posthoc one-way ANOVA showed a marginally significant effect for intrinsic vowel duration contrast ($F_{2, 277} = 2.93, p = 0.055$), where words with intrinsic vowel duration opposite to the lexical stress contrast (perception, permission) had somewhat longer WSD than words with symmetric vowel duration for both syllables.

Overlap among diagnostic groups.

For diagnostic purposes, it is important to examine not only differences in group distributions, but also the degree of individual overlap between and among diagnostic groups. As illustrated in Figure 1, degree of overlap between groups was greatest for the PVI_{dur} metric and smallest for the WSD metric. To evaluate this overlap further, we used a test of non-overlap developed for time series data; the Nonoverlap of All Pairs (Parker, Vannest, & Davis, 2011). Results, displayed in table 5, showed that although the AOS speakers, on average, differed from both APH and NTC speakers on all three duration-based metrics, the degree of nonoverlap was smallest for WSD (99.5% and 98.3% compared to APH and NTC groups, respectively). By comparison, percent nonoverlapping data ranged from 83.5% to 89.6% for PVI_{dur} and LSR.

Consistent with the post hoc analyses for group differences, there was also a significant proportion of nonoverlapping data between the APH group and the NTC group on both lexical stress measures, indicating that performance between these groups was not equivalent as would be the case if no prosodic impairment was present in the APH group. In contrast, the overlap between the APH and NTC groups was nearly complete for the WSD measure.

Discussion

Study results indicate that the WSD differentiated most effectively between AOS and aphasia without AOS and that PVI and LSR metrics were affected by the congruence between lexical stress assignment and intrinsic vowel duration.

The word syllable duration (WSD) was most informative about diagnosis

In answer to our first research question and consistent with previous studies that have reported prosodic impairment primarily in the temporal domain (Danly & Shapiro, 1982; Marquardt et al., 1995; Vergis et al., 2014), we conclude that of the five measures of word-

level prosody we investigated, AOS speakers differed significantly from neurotypical speakers and speakers with aphasia without AOS on metrics derived from duration measures. Of the three duration-based metrics, the most effective differentiation was seen with the WSD. This metric was also generated with the strongest interobserver agreement and highest specificity for AOS, given that speakers with aphasia and no AOS were indistinguishable from neurotypical controls. It will be important to confirm these results in larger samples. Because the WSD metric has the additional advantage of minimal segmentation requirements and observer judgement, clinical feasibility is promising and full automation is realistic.

The PVI_{dur} and LSR data were similar in terms of degree of overlap. Both metrics differentiated AOS from the other two groups, but both also yielded a significantly smaller stress contrast for the APH group than for the neurotypical control group. The primarily positive PVI_{dur} values indicate that all speaker groups likely used duration variation to mark differences in lexical stress. Further, the significantly lower values in the AOS group suggests that stress contrast, on average, had lower magnitude in this group. However, as illustrated in the Figure 1 box plots and the sums of squares listed in table 4, token-to-token PVI variability was substantial in all three groups. This variability may have indicated that speakers' attempts to produce our word targets were implemented with varying degrees of stress contrast or with varying relationships among the three dimensions known to mark stress. In a larger study, perceptual phonetic evaluation of lexical stress clarity may help differentiate between these explanations. Such exploration is also motivated by observations that lexical stress can be preserved in AOS (Marquardt et al., 1995; Odell & Shriberg, 2001; Skenes, 1987). Diagnostic efforts will benefit from both qualitative and quantitative information about the balance between preserved and impaired communication of lexical stress in speakers with AOS and how this balance compares to that of the disorder's diagnostic neighbors.

We note that inter-rater reliability for derived metrics, especially PVI_{dur} (ICC = 0.65), was lower than observed for the constituent acoustic measurement (e.g. nucleus duration; ICC = 0.90). This is predictable based on mathematical principles of error propagation for equations involving multiplication or division, including PVI and LSR metrics. In particular, when multiplying or dividing measured quantities, the error for the combined measure generally consists of the sum of the relative error in the constituent measures (Taylor, 1997 pp. 51–53), while error for a single measure will simply be the absolute measurement error for that quantity. This general principle of error propagation in equations accentuates the importance of reporting reliability estimates in research to match the dependent variables under investigation. It also indicates that additional attention to measurement accuracy may be necessary for clinical application of composite acoustic metrics. Target word or dialect-specific qualities, such as the rhotic vowel nuclei in the present study, may complicate segmentation and require more elaborate operational definitions.

It will be important to replicate these preliminary results in a larger participant sample and using phonetically balanced word targets. If the diagnostic sensitivity of WSD is replicated or PVI_{dur} stability improves in a homogeneous speech sample, meaningful clinical application will require normative data about cutoff scores based on specific speech targets.

Despite the obvious contributions of durational prosodic measures to AOS diagnosis, it is essential to remember that AOS is a multidimensional and behaviourally diagnosed syndrome with other prominent and potentially dissociable features in addition to abnormal word prosody. WSD, PVIDur, or any single measure, should not be considered pathognomonic. Other speech features, such as perceptible sound distortions and distorted substitution or addition errors, are also important for diagnosis.

Intrinsic vowel duration affected lexical stress metrics

Lexical stress was not the only factor that determined degree of duration contrast between the two syllables that were analysed in each word. PVIDur and LSR magnitude were significantly affected by the difference between the same syllables in terms of intrinsic vowel duration. Results supported the predictions we made based on observed synergy of the lexical stress contrast and intrinsic vowel duration relationship. When these variables were congruent, PVIDur was magnified. When they fell in opposite directions, PVIDur was potentially neutralized, resulting in smaller, or even negative, values. When the intrinsic vowel duration was approximately the same for the two syllables it did not directly influence PVIDur, which therefore assumed intermediate values. In comparison, WSD was predictably robust to target word vowel quality.

The observation that the intrinsic vowel duration contrast effects for PVIDur and LSR were most prominent for neurotypical control speakers, followed by APH speakers, can be interpreted as relative preservation of lexical stress and vowel quality in these groups. In contrast, for AOS speakers, who produced errors in both vowel quality and duration, none of the metrics showed a systematic effect for intrinsic vowel duration contrast. We did not code or otherwise measure vowel quality for the words we analyzed in this study. However, narrow phonetic transcription indicates that speakers with AOS produce numerous relatively subtle vowel errors (Haley, Bays, & Ohde, 2001; Haley et al., 2017; Odell et al., 1991). The causality between vowel quality and vowel duration may be interpreted in either direction. On the one hand, the production of segmental vowel errors may have altered intrinsic vowel duration and therefore reduced the duration contrast of the target vowel pair. On the other hand, if the AOS speakers produced syllable nuclei as prolongations, vowels with normally short intrinsic duration may give listeners the impression of vowel distortion or an altogether different phoneme (Klatt, 1976).

Future directions

The study results confirmed the importance of temporal variables to AOS diagnosis and demonstrated that important distribution and context sensitivity differences exist among quantification strategies. In light of the slow speaking rate associated with AOS and varying degrees of prolongation of both pauses and inter-syllabic segments, it is not surprising that any number of acoustic measures derived from these segments would show group differences relative to neurotypical speakers and speakers with aphasia and no AOS. By refraining from overinterpreting these group differences and instead proceeding with an eye toward psychometric quality improvement, we have an opportunity to advance to differential diagnostic practices in tangible and consequential ways.

Future research should also establish similarities and differences in how prosodic features of AOS distribute at the word level relative to longer linguistic units. As we remarked in the introduction, there are strong diagnostic reasons for quantifying prosody for repeated multisyllabic words. It is also true that the greater contextual complexity of word strings, sentences, and continuous speech can increase the magnitude of prosodic metrics such as those we used in this study (Strand & McNeil, 1996; Vergis et al., 2014). For some purposes, this difference may be advantageous and could even increase measurement and categorization precision.

Acknowledgments:

Gratitude is extended to our research assistants Michelle Swanson, Michael Smith, and Gabrielle Spinella, who helped with data analysis.

The project was supported by the NIH under grant R03DC006163.

References

- Arciuli J, & Ballard KJ (2017). Still not adult-like: lexical stress contrastivity in word productions of eight- to eleven-year-olds. *Journal of Child Language*, 44(5), 1274–1288. [PubMed: 27766991]
- Arciuli J, & Colombo L (2016). An acoustic investigation of the developmental trajectory of lexical stress contrastivity in Italian. *Speech Communication*, 80, 22–33.
- Ballard KJ, Djaja D, Arciuli J, James DGH, & van Doorn J (2012). Developmental trajectory for production of prosody: lexical stress contrastivity in children ages 3 to 7 years and in adults. *Journal of Speech, Language, and Hearing Research*, 55(6), 1822–1835.
- Ballard KJ, Wambaugh JL, Duffy JR, Layfield C, Maas E, Mauszycki S, & McNeil MR (2015). Treatment for acquired apraxia of speech: A systematic review of intervention research between 2004 and 2012. *American Journal of Speech-Language Pathology / American Speech-Language-Hearing Association*, 24(2), 316–337.
- Ballard Kirrie J, Azizi L, Duffy JR, McNeil MR, Halaki M, O'Dwyer N, ... Robin DA (2016). A predictive model for diagnosing stroke-related apraxia of speech. *Neuropsychologia*, 81, 129–139. [PubMed: 26707715]
- Ballard Kirrie J., Robin DA, McCabe P, & McDonald J (2010). A treatment for dysprosody in childhood apraxia of speech. *Journal of Speech, Language, and Hearing Research*, 53(5), 1227–1245.
- Ballard Kirrie J, Savage S, Leyton CE, Vogel AP, Hornberger M, & Hodges JR (2014). Logopenic and nonfluent variants of primary progressive aphasia are differentiated by acoustic measures of speech production. *Plos One*, 9(2), e89864. [PubMed: 24587083]
- Ballard Kirrie J, Wambaugh JL, Duffy JR, Layfield C, Maas E, Mauszycki S, & McNeil MR (2015). Treatment for acquired apraxia of speech: a systematic review of intervention research between 2004 and 2012. *American Journal of Speech-Language Pathology / American Speech-Language-Hearing Association*, 24(2), 316–337.
- Baum SR (1992). The influence of word length on syllable duration in aphasia: Acoustic analyses. *Aphasiology*, 6(5), 501–513.
- Boersma P, & Weenink D (2017). Praat: Doing phonetics by computer (6.0.31). Computer software
- Collins M, Rosenbek JC, & Wertz RT (1983). Spectrographic analysis of vowel and word duration in apraxia of speech. *Journal of Speech and Hearing Research*, 26(2), 224–230. [PubMed: 6887809]
- Cunningham KT, Haley KL, & Jacks A (2016). Speech sound distortions in aphasia and apraxia of speech: reliability and diagnostic significance. *Aphasiology*, 30(4), 396–413.
- Danly M, & Shapiro B (1982). Speech prosody in Broca's aphasia. *Brain and Language*, 16(2), 171–190. [PubMed: 7116126]
- Darley FL (1968). Apraxia of speech: 107 years of terminological confusion. Presented at the Annual meeting of the American Speech and Hearing Association, Denver, CO.

- Duffy J R. (2013). Motor speech disorders: substrates, differential diagnosis, and management (3rd ed.). St. Louis, MO: Mosby.
- Duffy Joseph R, Hanley H, Utianski R, Clark H, Strand E, Josephs KA, & Whitwell JL (2017). Temporal acoustic measures distinguish primary progressive apraxia of speech from primary progressive aphasia. *Brain and Language*, 168, 84–94. [PubMed: 28187331]
- Duffy Joseph R, Strand EA, Clark H, Machulda M, Whitwell JL, & Josephs KA (2015). Primary progressive apraxia of speech: clinical features and acoustic and neurologic correlates. *American Journal of Speech-Language Pathology / American Speech-Language-Hearing Association*, 24(2), 88–100.
- Haley KL, & Overton HB (2001). Word length and vowel duration in apraxia of speech: the use of relative measures. *Brain and Language*, 79(3), 397–406. [PubMed: 11781050]
- Haley KL (2002). Temporal and spectral properties of voiceless fricatives in aphasia and apraxia of speech. *Aphasiology*, 16(4–6), 595–607.
- Haley KL, Bays GL, & Ohde RN (2001). Phonetic properties of aphasic-apraxic speech: a modified narrow transcription analysis. *Aphasiology*, 15(12), 1125–1142.
- Haley KL, Jacks A, & Cunningham KT (2013). Error variability and the differentiation between apraxia of speech and aphasia with phonemic paraphasia. *Journal of Speech, Language, and Hearing Research*, 56(3), 891–905.
- Haley KL, Jacks A, de Riesthal M, Abou-Khalil R, & Roth HL (2012). Toward a quantitative basis for assessment and diagnosis of apraxia of speech. *Journal of Speech, Language, and Hearing Research*, 55(5), S1502–S1517
- Haley KL, Jacks A, Richardson JD, & Wambaugh JL (2017). Perceptually salient sound distortions and apraxia of speech: a performance continuum. *American Journal of Speech-Language Pathology*, 26, 631–640. [PubMed: 28654944]
- Haley KL, Roth H, Grindstaff E, & Jacks A (2011). Computer-mediated assessment of intelligibility in aphasia and apraxia of speech. *Aphasiology*, 25(12), 1600–1620. [PubMed: 22215933]
- Hallgren KA (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in quantitative methods for psychology*, 8(1), 23–34. [PubMed: 22833776]
- Helm-Estabrooks N (1992). *Aphasia diagnostic profiles* Austin, TX: Pro-Ed.
- Hillenbrand J, Getty LA, Clark MJ, & Wheeler K (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, 97(5), 3099–3111. [PubMed: 7759650]
- Kent RD, & Rosenbek JC (1982). Prosodic disturbance and neurologic lesion. *Brain and Language*, 15(2), 259–291. [PubMed: 7074345]
- Kent RD, & Rosenbek JC (1983). Acoustic patterns of apraxia of speech. *Journal of Speech and Hearing Research*, 26(2), 231. [PubMed: 6887810]
- Klatt DH (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 59(5), 1208–1221. [PubMed: 956516]
- Knight R-A, & Cocks N (2007). Rhythm in the speech of a person with right hemisphere damage: Applying the pairwise variability index. *Advances in Speech Language Pathology*, 9(3), 256–264.
- Koo TK, & Li MY (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163. [PubMed: 27330520]
- Lehiste I (1972). The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America*, 51, 2018–2024.
- Low EL, Grabe E, & Nolan F (2000). Quantitative characterizations of speech rhythm: syllable-timing in Singapore English. *Language and Speech*, 43(4), 377–401. [PubMed: 11419223]
- Marquardt TP, Duffy G, & Cannito MP (1995). Acoustic analysis of accurate word stress patterning in patients with apraxia of speech and broca's aphasia. *American Journal of Speech-Language Pathology*, 4(4), 180.
- Maruthy S, Venugopal S, & Parakh P (2017). Speech rhythm in Kannada speaking adults who stutter. *International Journal of Speech-Language Pathology*, 19(5), 529–537. [PubMed: 27576027]

- McNeil M, Robin D, & Schmidt R (2009). Apraxia of speech: definition, differentiation, and treatment. In McNeil MR (Ed.), *Clinical management of sensorimotor speech disorders* (2nd ed., pp. 249–268).
- Odell K, McNeil MR, Rosenbek JC, & Hunter L (1990). Perceptual characteristics of consonant production by apraxic speakers. *The Journal of Speech and Hearing Disorders*, 55(2), 345–359. [PubMed: 2329797]
- Odell Katharine, McNeil MR, Rosenbek JC, & Hunter L (1991). Perceptual characteristics of vowel and prosody production in apraxic, aphasic, and dysarthric speakers. *Journal of Speech and Hearing Research*, 34(1), 67–80. [PubMed: 2008083]
- Odell KH, & Shriberg LD (2001). Prosody-voice characteristics of children and adults with apraxia of speech. *Clinical Linguistics & Phonetics*, 15(4), 275–307.
- Parker RI, Vannest KJ, & Davis JL (2011). Effect size in single-case research: a review of nine nonoverlap techniques. *Behavior Modification*, 35(4), 303–322. [PubMed: 21411481]
- Seddoh SAK, Robin DA, Sim H-S, Hageman C, Moon JB, & Folkins JW (1996). Speech timing in apraxia of speech versus conduction aphasia. *Journal of Speech and Hearing Research*, 39(3), 590–603. [PubMed: 8783137]
- Shriberg LD, Aram DM, & Kwiatkowski J (1997a). Developmental apraxia of speech: II. toward a diagnostic marker. *Journal of Speech, Language, and Hearing Research*, 40 (2), 286–312.
- Shriberg LD, Aram DM, & Kwiatkowski J (1997b). Developmental apraxia of speech: III. a subtype marked by inappropriate stress. *Journal of Speech, Language, and Hearing Research*, 40(2), 313–337.
- Shriberg LD, Campbell TF, Karlsson HB, Brown RL, Mcsweeny JL, & Nadler CJ (2003). A diagnostic marker for childhood apraxia of speech: the lexical stress ratio. *Clinical Linguistics & Phonetics*, 17(7), 549–574. [PubMed: 14608799]
- Skenes LL (1987). Durational changes of apraxic speakers. *Journal of Communication Disorders*, 20(1), 61–71. [PubMed: 3819004]
- Smith MT, Haley KL, & Wambaugh JL (2017). Phonemic error analysis: edit distance instead of phonemic accuracy. poster presented at the annual convention of the American Speech-Language Hearing Association 11 9–11 Los Angeles, CA.
- Strand EA, & McNeil MR (1996). Effects of length and linguistic complexity on temporal acoustic measures in apraxia of speech. *Journal of Speech and Hearing Research*, 39(5), 1018–1033. [PubMed: 8898255]
- Taylor JR (1997). *An introduction to error analysis* 2nd ed. University Science Books: Sausalito, CA.
- van Santen JPH, Prud'hommeaux ET, & Black LM (2009). Automated assessment of prosody production. *Speech Communication*, 51(11), 1082–1097. [PubMed: 20160984]
- Ventry I, & Weinstein B (1983). Identification of elderly people with hearing problems. *ASHA*, 25, 37–42.
- Vergis MK, Ballard KJ, Duffy JR, McNeil MR, Scholl D, & Layfield C (2014). An acoustic measure of lexical stress differentiates aphasia and aphasia plus apraxia of speech after stroke. *Aphasiology*, 28(5), 554–575.
- Wertz RT, LaPointe LL, & Rosenbek JC (1984). *Apraxia of speech in adults: the disorder and its management* Orlando, FL: Grune & Stratton.
- Whitwell JL, Weigand SD, Duffy JR, Clark HM, Strand EA, Machulda MM, ... Josephs KA (2017). Predicting clinical decline in progressive agrammatic aphasia and apraxia of speech. *Neurology*, 89(22), 2271–2279. [PubMed: 29093069]

Figure 1.

Box plots illustrating the distribution of word production metrics for the three prosodic measure that were exclusively or primarily based on duration (PVI_{dur}, LSR, WSD). Units of the PVI_{dur} and LSR are defined by the ratio; WSD is expressed in seconds. Values are separated by speaker groups (AOS, APH, NTC) and pattern of intrinsic vowel duration (congruent with word stress, nonvarying with effects only for lexical stress, opposite to word stress). For each boxplot, the horizontal line represents the median, the ends of the box correspond to the 25th and 75th percentiles, and the whiskers represent the minimum and

maximum values in the range or, if there are outliers, the whiskers represent data points falling within 1.5 times the interquartile range of the 25th and 75th percentile.

Table 1.

Mean age and results of speech evaluation (SD in parenthesis).

Variable	AOS, n=7	APH, n=9	NTC, n=19
Age (years)	59.3 (17.0)	65.4 (9.8)	63.6 (12.5)
Aphasia Severity (percentile) ^a	45.9 (16.9)	75.8 (15.9)	NA
Phonemic edit distance ratio	25.7 (10.4)	6.5 (6.1)	NA
Distorted segments (%) ^b	15.7 (4.8)	10.1 (5.0)	NA
Single word intelligibility ^c	68.3 (8.5)	86.3 (5.9)	96.0 (3.4)
Multisyllabic WSD (ms) ^d	428.2 (160.1)	236.6 (36.6)	232.0 (31.1)
Single word PVIDur ^e	11.5 (60.1)	30.2 (28.2)	113.0 (15.5)

Note:

^aFrom the Aphasia Diagnostic Profiles

^bPercent phonemes on the motor speech evaluation that were produced with one or more distortion errors, as determined by narrow phonetic transcription.

^cFrom the CHMIT-e

^dFrom six multisyllabic words produced during the motor speech evaluation.

^eFrom nucleus duration data for five repetitions of the word 'catastrophe'.

Data were not available on this measure for four speakers (three in the AOS group, one in the APH group) who declined repeating this word. NA=Not Applicable, because measures were not derived for the neurotypical control group.

Table 2.

Syllable comparisons and anticipated duration difference for the nine target words, based on lexical stress contrast and intrinsic nucleus duration difference

Target word	Syllable pairs	Intrinsic Vowel Duration (unstressed – stressed)	Vowel Duration Change Pattern
combination	2 and 3	short–long	congruent
publication	2 and 3	short–long	congruent
conversation	2 and 3	long–long	neutral
collection	1 and 2	short–short	neutral
condition	1 and 2	short–short	neutral
connection	1 and 2	short–short	neutral
position	1 and 2	short–short	neutral
perception	1 and 2	long–short	opposite
permission	1 and 2	long–short	opposite

Table 3.

Reliability of prosodic measures

	Syllable nucleus measures			Word-level measures				
	Nucleus Duration	Intensity	F ₀	PVIdur	PVInt	PVIfreq	LSR	WSD
ICC	0.90	0.98	0.99	0.65	0.70	0.96	0.67	0.99

Note: Intraclass correlations (ICC) were computed using a two-way mixed effects model with random rater effects and fixed measures effects; coefficients represent absolute agreement. F₀=fundamental frequency. PVIdur=Pairwise variability index for duration, PVInt=Pairwise variability index for intensity, PVIfreq=Pairwise variability index for fundamental frequency, LSR=Lexical stress ratio, WSD=Word Syllable Duration.

Table 4.

Results of 2-way ANOVAs for each of the five prosodic measures

Metric and factor	df	SS	F Ratio	p
PVI _{dur}				
Speaker group	2	182624.65	73.73	<.0001*
Intrinsic vowel duration	2	4694.74	26.12	<.0001*
Speaker group x intrinsic vowel duration	4	17927.66	3.62	.0068*
PVI _{int}				
Speaker group	2	16.40	0.31	0.7330
Intrinsic vowel duration	2	43.72	0.83	0.4376
Speaker group x intrinsic vowel duration	4	35.52	0.34	0.8531
PVI _{frq}				
Speaker group	2	1405.23	2.45	0.0885
Intrinsic vowel duration	2	2537.42	4.10	0.0175*
Speaker group x intrinsic vowel duration	4	2182.63	1.90	0.1107
LSR				
Speaker group	2	29.88	85.80	<.0001*
Intrinsic vowel duration	2	8.57	24.60	<.0001*
Speaker group x intrinsic vowel duration	4	2.18	3.13	.0153*
WSD				
Speaker group	2	1.66	183.04	<.0001*
Intrinsic vowel duration	2	0.03	2.93	.0553
Speaker group x intrinsic vowel duration	4	0.03	1.83	.1241

Note: PVI_{dur}=Pairwise variability index for duration, PVI_{int}=Pairwise variability index for intensity, PVI_{frq}=Pairwise variability index for fundamental frequency, LSR=Lexical stress ratio, WSD=Word Syllable Duration.

Table 5.

Results of Nonoverlap of All Pairs (NAP) analyses for the three prosodic measures that differed across groups. Values are expressed in the direction of the mean/median difference, such that 0.5 indicates complete overlap and 1.0 indicates no overlap. Significance values indicate the probability that a randomly chosen measure in one group will be greater (or lesser) than one chosen from the other group.

	PVIdur	LSR	WSD
AOS vs NTC	0.8958 <i>p</i> <0.0001	0.8962 <i>p</i> <0.0001	0.9951 <i>p</i> <0.0001
AOS vs APH	0.8480 <i>p</i> <0.0001	0.8345 <i>p</i> <0.0001	0.9827 <i>p</i> <0.0001
NTC vs APH	0.6067 <i>p</i> =0.0078	0.6177 <i>p</i> =0.0034	0.5406 <i>p</i> =0.3123

Note: PVIdur=Pairwise variability index for duration of syllable nuclei, LSR = Lexical stress ratio, WSD = word syllable duration.