Research Article

A Multivariate Analytic Approach to the Differential Diagnosis of Apraxia of Speech

Alexandra Basilakos,[a] Grigori Yourganov,[b] Dirk-Bart den Ouden,[a] Daniel Fogerty,[a] Chris Rorden,[b,c] Lynda Feenaughty,[a,d] and Julius Fridriksson[a,c]

Purpose: Apraxia of speech (AOS) is a consequence of stroke that frequently co-occurs with aphasia. Its study is limited by difficulties with its perceptual evaluation and dissociation from co-occurring impairments. This study examined the classification accuracy of several acoustic measures for the differential diagnosis of AOS in a sample of stroke survivors.
Method: Fifty-seven individuals were included (mean age = 60.8 ± 10.4 years; 21 women, 36 men; mean months poststroke = 54.7 ± 46). Participants were grouped on the basis of speech/language testing as follows: AOS-Aphasia (*n* = 20), Aphasia Only (*n* = 24), and Stroke Control (*n* = 13). Normalized Pairwise Variability Index, proportion of distortion errors, voice onset time variability, and amplitude envelope modulation spectrum variables

were obtained from connected speech samples. Measures were analyzed for group differences and entered into a linear discriminant analysis to predict diagnostic classification.
Results: Out-of-sample classification accuracy of all measures was over 90%. The envelope modulation spectrum variables had the greatest impact on classification when all measures were analyzed together.
Conclusions: This study contributes to efforts to identify objective acoustic measures that can facilitate the differential diagnosis of AOS. Results suggest that further study of these measures is warranted to determine the best predictors of AOS diagnosis.
Supplemental Materials: https://doi.org/10.23641/asha.5611309

Apraxia of speech (AOS) is a motor speech disorder characterized by impaired planning and programming of articulatory movements. A large body of research has identified specific speech characteristics consistent with AOS (e.g., Ballard et al., 2016; Cunningham, Haley, & Jacks, 2016; Galluzzi, Bureca, Guariglia, & Romani, 2015; Haley, Jacks, & Cunningham, 2013; Jacks, Mathes, & Marquardt, 2010; Kent & Rosenbek, 1983; McNeil, Robin, & Schmidt, 1997; Odell, McNeil, Rosenbek, & Hunter, 1990; Odell, McNeil, Rosenbek, & Hunter, 1991; Ogar et al., 2006; Rosenbek, Kent, & Lapointe, 1984; Varley & Whiteside, 2001; Vergis et al., 2014; Wertz,

LaPointe, & Rosenbek, 1984; Ziegler & von Cramon, 1985, 1986a, 1986b). However, diagnosing AOS is challenging due to its frequent co-occurrence with aphasia (a higher level linguistic impairment) and dysarthria (an impairment in speech execution and control). Speech production is impaired in different ways in AOS, dysarthria, and aphasia, but perceptually, speech sound errors that occur in all three disorders can manifest similarly. Despite attempts to improve characterization of apraxic behaviors, there is no universally accepted definition or highly reliable "gold standard" for its diagnosis (McNeil, Pratt, & Fossett, 2004; Mumby, Bowen, & Hesketh, 2007).

This study is an investigation into several acoustic measures for the differential diagnosis of AOS. Although differential diagnosis between AOS and dysarthria is problematic, this study specifically focuses on the differential diagnosis of AOS from poststroke aphasia. The following sections will review literature on the assessment and diagnosis of AOS, followed by a detailed description of the measures utilized in this study.

Overview of Diagnostic Criteria

In clinical practice, speech-language pathologists (SLPs) often rely upon perceptual judgment to evaluate a patient's

[a]Department of Communication Sciences & Disorders, University of South Carolina, Columbia
[b]Department of Psychology, University of South Carolina, Columbia
[c]McCausland Center for Brain Imaging, University of South Carolina, Columbia
[d]Department of Neurology, Medical University of South Carolina, Charleston

Correspondence to Alexandra Basilakos: basilaka@mailbox.sc.edu

speech, but perceptual evaluation is subject to the pitfalls of categorical perception (Buckingham & Yule, 1987; Code, 1998) and may suffer from reduced interrater reliability (e.g., Haley, Jacks, de Riesthal, Abou-Khalil, & Roth, 2012). The same problems can contaminate diagnostic labeling in research studies, especially if research groups adopt their own diagnostic criteria (or interpretations of these criteria), which may not be uniform across sites (McNeil et al., 1997; Mumby et al., 2007). As discussed extensively elsewhere, these issues constitute a major challenge for the clinical management of AOS (Ballard, Granier, & Robin, 2000; Haley et al., 2012; McNeil et al., 2004; Wambaugh, Wright, Nessler, & Mauszycki, 2014).

Early objective analyses characterized AOS with acoustic (e.g., Collins, Rosenbek, & Wertz, 1983; Kent & Rosenbek, 1983; McNeil, Liss, Tseng, & Kent, 1990; Seddoh et al., 1996; Square-Storer & Appledoorn, 1991) and electromyographic/kinematic measures (e.g., Fromm, Abbs, McNeil, & Rosenbek, 1982; Itoh & Sasanuma, 1984; Itoh, Sasanuma, & Ushijima, 1979; McNeil & Adams, 1991; Robin, Bean, & Folkins, 1989; Shankweiler & Harris, 1966). These studies provided more objective characterizations of speech and motor performance and influenced the development of clinical diagnostic markers (e.g., for discussion: Rosenbek et al., 1984; Wertz et al., 1984). For example, Wertz et al. (1984) published the earliest and perhaps most widely utilized diagnostic description of AOS. This description is based on work from Darley, Aronson, and Brown (1975) and from studies on the behavioral and acoustic characteristics of AOS (e.g., Rosenbek et al., 1984; Wertz et al., 1984). Wertz et al. (1984) define AOS by its articulatory struggle and "effortful groping," multiple (often failed) attempts at self-correction, inconsistency in repeated productions, and dysprosody (described as atypical rhythm, stress, and intonation). Later, the Apraxia Battery for Adults (ABA) was published (Dabul, 1979) and revised (Apraxia Battery for Adults–Second Edition [ABA-2]; Dabul, 2000) and, for over three decades, was one of the only available psychometric assessments for AOS. Critiques of the ABA-2 argue that it is limited in the extent that it provides a reliable differential diagnosis between AOS and conduction aphasia (McNeil et al., 2004). For example, in some subtests, speech production is assessed from repetition tasks and scored according to production delays, variability, or self-corrections—errors that are not unique to AOS, but may also occur in aphasia. Most recently, the Apraxia of Speech Rating Scale (ASRS; Strand, Duffy, Clark, & Josephs, 2014) was developed to rate the presence and severity of speech sound disorders that can occur uniquely in AOS, in both AOS and/or aphasia, AOS and/or dysarthria, or all three disorders. The features highlighted as those that occur uniquely in AOS include distorted sound substitutions, distorted sound additions, increased frequency of distortion errors upon increased articulatory length and complexity, off-target sequential motor rates, and reduced phrase length not due to inadequate breath support (Strand et al., 2014). Preliminary validation of the ASRS indicates that it has

good reliability and validity (see Strand et al., 2014 for psychometric properties).

Although these diagnostic criteria were developed from extensive research, the abovementioned rating scales and assessments essentially rely on clinicians' perceptual evaluations, which is a limitation given that subtle motor abnormalities may go unnoticed by listeners (Fromm et al., 1982). Recent work has sought to decrease subjectivity in AOS diagnosis by studying the extent that more objective acoustic measures may be viable diagnostic tools to supplement clinicians' perception of speech (Duffy, 2005). These studies have investigated measures of prosody (Ballard et al., 2016; Vergis et al., 2014), articulatory imprecision (Cunningham et al., 2016), or a combination of these measures alongside other cognitive–linguistic and demographic variables (e.g., Ballard et al., 2016). Collectively, this work has shown that individuals with AOS (and concomitant aphasia) differ from those without AOS (either with aphasia or without any speech/language impairment) on measures of prosody and the occurrence of sound-level distortion errors (Ballard et al., 2016; Cunningham et al., 2016; Vergis et al., 2014). However, apart from Ballard et al.'s study (2016), the aforementioned studies that have objectively quantified prosody (e.g., Vergis et al., 2014) and distortion errors (e.g., Cunningham et al., 2016) have included relatively small sample sizes, limiting the extent to which results can be generalized to the clinical population and across research sites. Therefore, the current study sought to replicate and extend prior work with an independent sample of stroke survivors. A brief overview and discussion of pertinent acoustic measures is provided in the sections that follow.

Specific Features of Apraxic Speech

Speech Prosody in AOS

The most recent studies to investigate prosodic disturbances in poststroke AOS (Ballard et al., 2016; Vergis et al., 2014) used the Pairwise Variability Index (PVI; Grabe & Low, 2002), a coefficient obtained by computing the relative difference between vowel durations in multisyllabic words, or across phrases and sentences. AOS has been characterized by a lack of syllabic contrastiveness due to equal and excess stress placement (Kent & Rosenbek, 1983), and the PVI coefficient provides a quantification of this behavioral impairment. In a small sample of participants with AOS and concomitant aphasia ($n = 9$), aphasia only ($n = 8$), and control individuals ($n = 8$), Vergis et al. (2014) found that the PVI measures for vowel duration (henceforth, Pairwise Variability Index–Vowels [PVI-V]) were similar for the individuals with aphasia and the control group. In contrast, the individuals with AOS and aphasia demonstrated a significantly smaller vowel duration contrast for words that followed a weak–strong stress pattern (e.g., p_o_t_a_t_o) when compared with those with aphasia only and control individuals. Ballard et al. (2016) replicated this finding in a larger sample ($N = 72$; albeit, in a sample that also included data from Vergis et al.'s participants). These

results suggest that PVI-V shows promise for the differential diagnosis of AOS (Ballard et al., 2016).

In the dysarthria literature, Liss, LeGendre, and Lotto (2010) showed that measures derived from the amplitude envelope modulation spectrum (EMS) may provide a more automated measure of speech prosody to differentially diagnose dysarthria variants, potentially with the same (or greater) accuracy as PVI or related coefficients (Liss et al., 2009, 2010). Speech is characterized by modulations in amplitude that correspond to different features of the speech signal, ranging from rhythmic fluctuations associated with prosody and syllabic nuclei to faster cycles associated with rapid articulatory movements (Crouzet & Ainsworth, 2001; Hall & Grose, 1993). Quantification of speech energy across the amplitude modulation spectrum can be used to investigate these aspects of production. For example, slower modulations, that is, at 1 Hz–2 Hz, have been associated with the prosodic contour of connected speech (Ghitza, 2011; Ghitza & Greenberg, 2009; Liss et al., 2010), and modulations within the 4 Hz–8 Hz range have been associated with the regularity of syllabic production, corresponding to cycles of jaw opening and closure (Giraud & Poeppel, 2012; MacNeilage, 1998). Faster modulations (i.e., >15 Hz–32 Hz) are reflective of the brief duration of phonetic features that occur in speech production, such as in the production of consonants (e.g., Ghitza, 2011; Giraud et al., 2007). Each of these levels of production—prosodic, syllabic, and phonetic—is commonly affected to some degree in AOS (Ballard et al., 2016; Cunningham et al., 2016; Vergis et al., 2014; Ziegler, 2005, 2009). Thus, EMS analysis may be useful in measuring these production differences in individuals with AOS across various speech timescales. To our knowledge, no published studies have investigated the utility of EMS measures in the differential diagnosis of AOS.

Phonetic Errors in AOS

A key debate in the AOS literature has focused on the extent to which apraxic errors are phonemic or phonetic in nature, even while it has long been acknowledged that the distinction between the two may not be so clearly dichotomous (Code, 1998). Several studies have attempted to classify speech errors into phonetic and phonemic categories. In cases where such classifications are based on phonetic transcriptions, narrow phonetic transcription should be preferred to broad transcription, as broad transcription does not fully capture distorted productions, potentially leading to an inflated estimate of phonemic errors (Canter, Trost, & Burns, 1985; Cunningham et al., 2016; Miller, 1995; Odell et al., 1990; Rosenbek et al., 1984). However, as reviewed by Cunningham et al. (2016), few studies have actually used narrow phonetic transcription to investigate the frequency of distorted productions, and within those studies, findings have been mixed (Canter et al., 1985; Miller, 1995).

In a comprehensive evaluation of articulation errors, Cunningham et al. (2016) categorized 15 speakers with focal left hemisphere injuries into two groups on the basis of the presence/absence of atypical prosody, quantified by a "word syllable duration" (WSD) measure. Those with longer WSD scores were assigned to a "probable AOS" group (P-AOS, given that atypical prosody is often associated with AOS), and those with normal WSD scores were assigned to a "probable aphasia with phonemic paraphasia" group (P-APP). Cunningham et al. (2016) found that individuals in the P-AOS group had more distortion errors than the P-APP group and that most errors could be categorized as consonant voicing errors (39.3%), segmental lengthening errors (27.7%), and errors with tongue placement (24.9%). Distortion errors were not unique to the P-AOS group, as the P-APP group demonstrated distortion errors on 2.8% of segments produced. Nevertheless, the authors conclude that the quantification of distortion errors has preliminary diagnostic value for AOS but that further research is needed regarding the frequency and type of distortion errors produced by individuals with AOS compared with those without. Currently, there is no normative reference for the use of distortion errors in the diagnosis of AOS. Taken together, the frequency of distortion errors indeed appears to be a valid diagnostic marker of AOS, but more work on this topic is warranted.

Voice onset time (VOT) has also been used to adjudicate between phonemic and phonetic errors in speakers with AOS. Voiced/Voiceless contrasts have been used to investigate whether substitution errors are indeed substitutions (e.g., substituting /p/ for /b/), or if subtle changes in articulation and voicing are perceived as such due to categorical perception. In one such study, Itoh et al. (1982) showed that individuals with AOS and conduction aphasia differed on the rate of phonetic errors produced (i.e., VOTs that fell outside the range expected for the given phoneme), with those with AOS demonstrating more phonetic errors than those with conduction aphasia (who did not differ from controls). Individuals with AOS (and concomitant aphasia) have also been shown to have more *variable* VOT than those with conduction aphasia (Seddoh et al., 1996). These studies support the use of VOT in obtaining information to characterize phonetic production beyond what could be provided from perceptual analysis of production.

Results from these studies provide empirical support regarding differences in prosody and articulation in speakers with AOS (and concomitant aphasia) compared with those with aphasia only. However, it remains uncertain which of these characteristics, or combination thereof, best distinguishes between speakers with and without AOS (see also Ballard et al., 2016 for discussion). Perhaps, the main reason for this uncertainty stems from the fact that there are few studies that have obtained a variety of speech production measures simultaneously from a large sample of speakers (for exceptions, see Ballard et al., 2016; Haley et al., 2012). Moreover, many measures can be time consuming to obtain from speech samples, especially from a large cohort of individuals. Therefore, the goal of this study was to provide additional data on acoustic measures that best discriminate between speakers with AOS and those

without by comparing several measures that have been reported in the AOS literature (i.e., normalized Pairwise Variability Index [nPVI], VOT, distortion errors). This study also serves as a preliminary investigation into EMS measures in AOS.

Method

Participants

This was a retrospective study that included data from individuals who completed speech/language testing and neuroimaging within the Aphasia Lab at the University of South Carolina over the past 10 years. The current sample was selected from a larger study that recruited individuals with a history of single-event ischemic stroke, in the chronic phase of recovery (more than 6 months poststroke), and between the ages of 20–80 years (both at time of stroke and testing). Only individuals with left hemisphere stroke were considered for this study. Exclusion criteria included history of neurological disorder affecting the brain (e.g., dementia), traumatic brain injury, or developmental speech–language impairment. Aphasia severity was evaluated using the Aphasia Quotient (AQ) composite score from the Western Aphasia Battery (WAB; Kertesz, 2006). Individuals who had a diagnosis of severe aphasia (i.e., WAB AQ score < 20) were excluded, as these participants' speech output was insufficient for analysis (i.e., too few words/utterances produced to analyze).

From the initial sample of 77 individuals who completed testing at the time of analysis, the final sample analyzed here included 57 participants (mean age = 60.8 ± 10.4 years; 21 women, 36 men; months postonset = 54.7 ± 46). All participants consented to study procedures by signing an informed consent form approved by the institutional review board at the University of South Carolina. All participants were reimbursed for their time and were compensated for travel.

Participants were assigned to one of three groups on the basis of their ASRS and WAB AQ scores. Note that because this study was retrospective, ASRS and WAB AQ scores were obtained from our participant database. The WAB was administered by American Speech-Language-Hearing Association (ASHA)–certified SLPs with extensive experience working with individuals with aphasia. The primary author (AB) and a secondary rater scored the archived ASRS for prior studies (see Basilakos, Rorden, Bonilha, Moser, & Fridriksson, 2015 and Moser, Basilakos, Fillmore, & Fridriksson, 2016 for additional details). According to the ASRS, 20 participants demonstrated behaviors consistent with AOS (i.e., an ASRS total score of ≥ 8, at least one behavior unique to AOS rated as present; Strand et al., 2014). Of these 20, all but two had a diagnosis of aphasia according to WAB diagnostic criteria (WAB AQ < 93.8); all were assigned to the AOS-Aphasia group. Note that because only two participants had AOS as their only impairment, they were included with the other participants in the AOS-Aphasia group. The remaining

38 participants without AOS could be further classified with aphasia only (n = 24 participants, Aphasia Only group) or no chronic poststroke communication impairment measured by WAB or ASRS criteria (n = 13, Stroke Control group). Group characteristics are presented in Table 1.

Behavioral Assessments

All speech production measures used in this study were obtained from three connected speech samples: the "cookie theft" picture from the Boston Diagnostic Aphasia Examination (Goodglass, Kaplan, & Barresi, 2000), the "circus" picture from the ABA-2 (Dabul, 2000), and the "picnic scene" from the WAB (Kertesz, 2006). Pictures were displayed on a Dell Vostro 3550 laptop computer, and participants were given 2 min per picture to describe it. All picture descriptions were audio–video-recorded using the same experimental computer for offline analysis. Participants were seated comfortably in an audiologic suite to attenuate background noise while completing experimental tasks. The mean number of syllables per second was as follows for each group: Stroke Control: 2.5 ± 0.43; Aphasia Only: 1.6 ± 0.81; and AOS-Aphasia: 0.8 ± 0.41.

Procedure

Audio–video-recorded picture descriptions were converted to WAV format files (44-kHz sampling rate) and annotated using Praat sound analysis software (Boersma & Weenink, 2001). An ASHA-certified SLP with experience in the assessment of individuals with poststroke speech and language deficits completed all transcription and analysis procedures (AB). Another ASHA-certified SLP with extensive experience with acoustic analysis served as a secondary rater for reliability (LF). Details of each measure are mentioned below.

PVI

To obtain nPVI-V coefficients, vocalic segments were identified in Praat (Boersma & Weenink, 2001). Segmentation was completed according to guidelines by Peterson and Lehiste (1960) with additional considerations for spontaneous speech (Thomas & Carter, 2006) and speakers with production disorders (Liss et al., 2009). All intervals were segmented according to visible formant structures (for vowel onset/offsets) and spectral energy corresponding to different consonant classes (for consonant onsets/offsets). Detailed segmentation criteria can be found in the Supplemental Material S1 (Figures S1a and S1b).

Once vowel durations were obtained, nPVI-V was calculated from the comparison of the duration of each successive vocalic interval using the following formula:

$$\text{nPVI} = 100 \times \left[\sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} \right| / (m-1) \right] \quad (1)$$

Table 1. Characteristics of the study sample.

Characteristic	Stroke Control (*n* = 13)	Aphasia Only (*n* = 24)	AOS-Aphasia (*n* = 20)	Significant differences*
Mean age	64.8 ± 10.6	61.4 ± 9.2	58.9 ± 10.5	SC = AO = A-A
Men:Women	4:9	19:5	13:7	
WAB AQ	98.4 ± 1.1	74.9 ± 20.2	64.2 ± 21.4	SC > AO > A-A
Aphasia types		Anomic: 11	Anomic: 4	
		Broca's: 3	Broca's: 13	
		Conduction: 7	Global: 1	
		Wernicke's: 3	None: 2[a]	
ASRS AOS severity	0	0	2.8 ± 1.0	SC = AO
				A-A > SC
				A-A > AO
ASRS aphasia severity	0	1.6 ± 1.0	2.2 ± 1.3	SC < AO
				SC < A-A
				AO = A-A
ASRS dysarthria severity	0	0.1 ± 0.9	0.5 ± 0.8	SC = AO
				SC < A-A
				AO = A-A

Note. AOS = Apraxia of Speech; SC = Stroke Control; AO = Aphasia Only; A-A = AOS-Aphasia; WAB AQ = Western Aphasia Battery Aphasia Quotient; ASRS = Apraxia of Speech Rating Scale.

[a]These two individuals were not aphasic according to WAB AQ cutoff criterion of 93.8.

*Based on independent samples *t* tests, *p* < .05.

where *d* is the duration of the selected interval, *k* is the chosen interval, and *m* is the number of vocalic intervals in which the PVI equation is calculated (Grabe & Low, 2002).

Articulatory Distortion Errors

Narrow transcription codes were adapted from Cunningham et al. (2016). Distorted productions resulting from incorrect tongue placement were identified and transcribed with diacritic markers. Distortion types included productions that were dentalized/fronted, palatized/backed, lateralized, rhotacized, derhotacized, and frictionalized. Distortions were summed across error type, and the proportion of distortion errors was calculated as the total number of distortions per words produced in the speech samples. Further details can be found in the Supplemental Material S2.

VOT

VOT was obtained for all word–initial plosives within each speech sample. The interval from the beginning of the burst release to the onset of voicing was demarcated in Praat. VOT variability was calculated separately for voiced and voiceless stop consonants as the standard deviation of VOT.

Amplitude EMS

To obtain the EMS measures, all audio recordings were first preprocessed using Adobe Audition CC to filter noise and remove extraneous sounds (e.g., clinician interjection). Details of the preprocessing steps are outlined in the Supplemental Material S3. Subsequent procedures were carried out in MATLAB using custom scripts. EMS was calculated from the first 90 s of each recording, over three 20-s windows at the beginning (0 s–20 s), middle (35 s–55 s), and end (70 s–90 s) of each sample. Prior EMS studies (e.g., Liss et al., 2010) analyzed shorter speech segments

(i.e., sentence-level stimuli); therefore, our window duration was chosen to ensure that a sufficient amount of speech was analyzed (Tilsen & Johnson, 2008).

For each 20-s interval, speech was downsampled to a frequency of 16 kHz. Envelopes were extracted via half-wave rectification, and the envelope was low-pass filtered using a sixth-order Butterworth filter at 50 Hz (Leong, Stone, Turner, & Goswami, 2014; Rosen, 1992). The low-pass filtered envelope was then downsampled to a sampling frequency of 1000 Hz. Next, the fast Fourier transform (FFT) was computed. Energy was summed into FFT bins that corresponded to octave bands with the center frequencies ranging from 1 Hz to 32 Hz. Finally, the energy in an octave-band FFT bin was divided by the energy in the 0 bin, yielding the normalized modulation index relative to the DC offset. Variables were derived from octave bands, selected to reflect the prosodic (1 Hz–2 Hz band), syllabic (4 Hz–8 Hz band), and articulatory (16 Hz–32 Hz band) aspects of speech. Octave bands were combined in this way because amplitude modulation spectra are conventionally reported for a given range (e.g., Giraud et al., 2007; Leong et al., 2014; Poeppel, 2003).

To obtain each of the amplitude modulation bands, amplitude energy was summed across octave bands and divided by the total energy in the 1 Hz–32 Hz spectrums. The resulting energy for each EMS band was then averaged across the three sample windows obtained for each participant. We chose to average across the three windows in attempt to obtain stable estimates of the EMS within the context of the typical variability in production that can occur in aphasia (e.g., Murray, Holland, & Beeson, 1998). It is important to note that there were no statistically significant differences in amplitude energy for the three EMS bands across any of the 20-s windows (Bonferroni-corrected *p* value < .006; i.e., corrected for nine total

comparisons for each group: three EMS bands over three time windows).

An example EMS plot for a typical speaker is presented in Figure 1.

Reliability

Reliability of the perceptual (distortion errors) and manual acoustic measures (VOT, vowel durations used for nPVI coefficients) was established using a two-way mixed consistency single-measure intraclass correlation coefficient (ICC). Speech samples from six individuals (10.3% of the study sample) were randomly selected, and the secondary rater (LF) was blind to ASRS scores, results from other testing, and the primary rater's scores. ICC values for each measure were as follows: distortion errors: .87; nPVI: .83; and mean VOT (collapsed across voiced and voiceless targets): .98. Because ASRS scores were obtained from the archival database, the primary rater (AB) randomly selected and scored an additional six speech samples from participants included in the current study to establish reliability pertinent to this work. ICC for the sum of all ASRS ratings = .90.

Intrarater reliability was established by randomly selecting six different participants. ICC values for intrarater reliability are as follows: distortion errors: .91; nPVI: .96; VOT: .96; ASRS (total of all ratings): .90: All ICC values are considered *good* to *excellent* (Cicchetti, 1994) and are in line with reliability measures reported in other AOS studies (e.g., Cunningham et al., 2016; Vergis et al., 2014).

Data Analyses

Data analyses were completed in two steps. First, all variables were inspected for group differences using an omnibus test suitable for their respective distributions—either univariate or multivariate analysis of variance (MANOVA;

Figure 1. Example EMS plot for a typical speaker. The top waveform depicts the raw audio file, the middle portion depicts the extracted amplitude envelope, and the bottom portion displays the modulation index for each modulation band. EMS = envelope modulation spectrum.
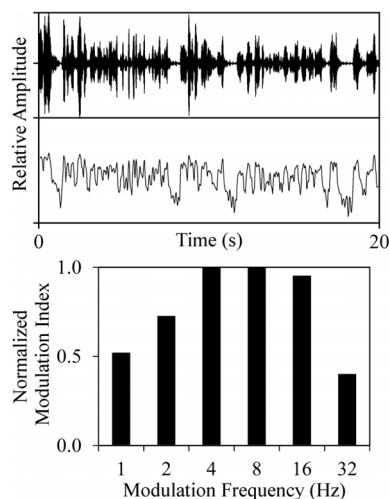
used for normally distributed variables) or a Kruskal–Wallis nonparametric analysis of variance (ANOVA; for non-normally distributed variables). Analyses were completed with SPSS, Version 24. For all models with a significant main effect of group, post hoc pairwise comparisons were conducted with *t* tests (normally distributed variables) or Dunn–Bonferroni post hoc tests (nonparametric tests; Dunn, 1964). All pairwise comparisons were Bonferroni corrected for multiple comparisons according to the default SPSS method. Details of each model and statistical assumptions are presented in the Results section.

Next, classification analyses were completed to determine the extent to which the experimental measures were predictive of AOS (using the ASRS criteria to define presence or absence of AOS). Classification analyses were then completed using Fisher's linear discriminant analysis (LDA; Fisher, 1936) to evaluate whether individuals with AOS can be distinguished from those without AOS on the basis of the measures investigated in this study. The LDA also provided information regarding the relevance of variables in classification. This analysis was completed in MATLAB using the Statistics and Machine Learning toolbox and custom scripts.

To be entered into the LDA, all measures were standardized to *z*-score distributions. Higher *z* scores were defined to correspond with better performance. Variables in the opposite direction (i.e., where a higher value corresponded to worse performance) were reversed by multiplying each *z* score value by −1. Importantly, the LDA implemented a leave-one-out cross-validation approach to avoid overfitting the model. This is an iterative process where one subject was set aside and the relationship between the behavioral variables and group membership was determined using the remaining *N*-1 subjects (where *N* is the original sample size). This relationship was used to predict the group of the set-aside subject; the subject was consequently put back into the pool, and another subject was set aside. The process was repeated until all subjects were, in turn, set aside and their group membership was predicted. Classification accuracy was computed as the percentage of subjects for whom the group membership was predicted correctly. Variables were entered into the LDA regardless of correlations; LDA can handle correlations between variables and uses the correlations to yield a more accurate separation of the classes (i.e., the presence/absence of AOS; for a general example, see Box 2 in Haynes & Rees, 2006). See Figure S2 (Supplemental Material S4) for correlation coefficients between each variable.

Results

Results by Variable

nPVI-V

Each group's nPVI-V distribution met the assumption of normality. nPVI-V scores for each group are displayed in Figure 2a as box plots, and mean nPVI-V (± *SD*) are as follows: Stroke Control: 62.94 ± 3.61; Aphasia

Figure 2. Box plots of the proportion of nPVI-V (Panel A), distortion errors (Panel B), and VOT variability for voiced (Panel C) and voiceless (Panel D) initial stop consonants. Horizontal lines indicate group differences at Bonferroni-corrected *p* level of .0167. Outliers are indicated by the *x*. SC = Stroke Control; AO = Aphasia Only; AOS-A = Apraxia of Speech with Concomitant Aphasia; nPVI-V = normalized Pairwise Variability Index–Vowels; VOT-SD = standard deviation of voice onset time.
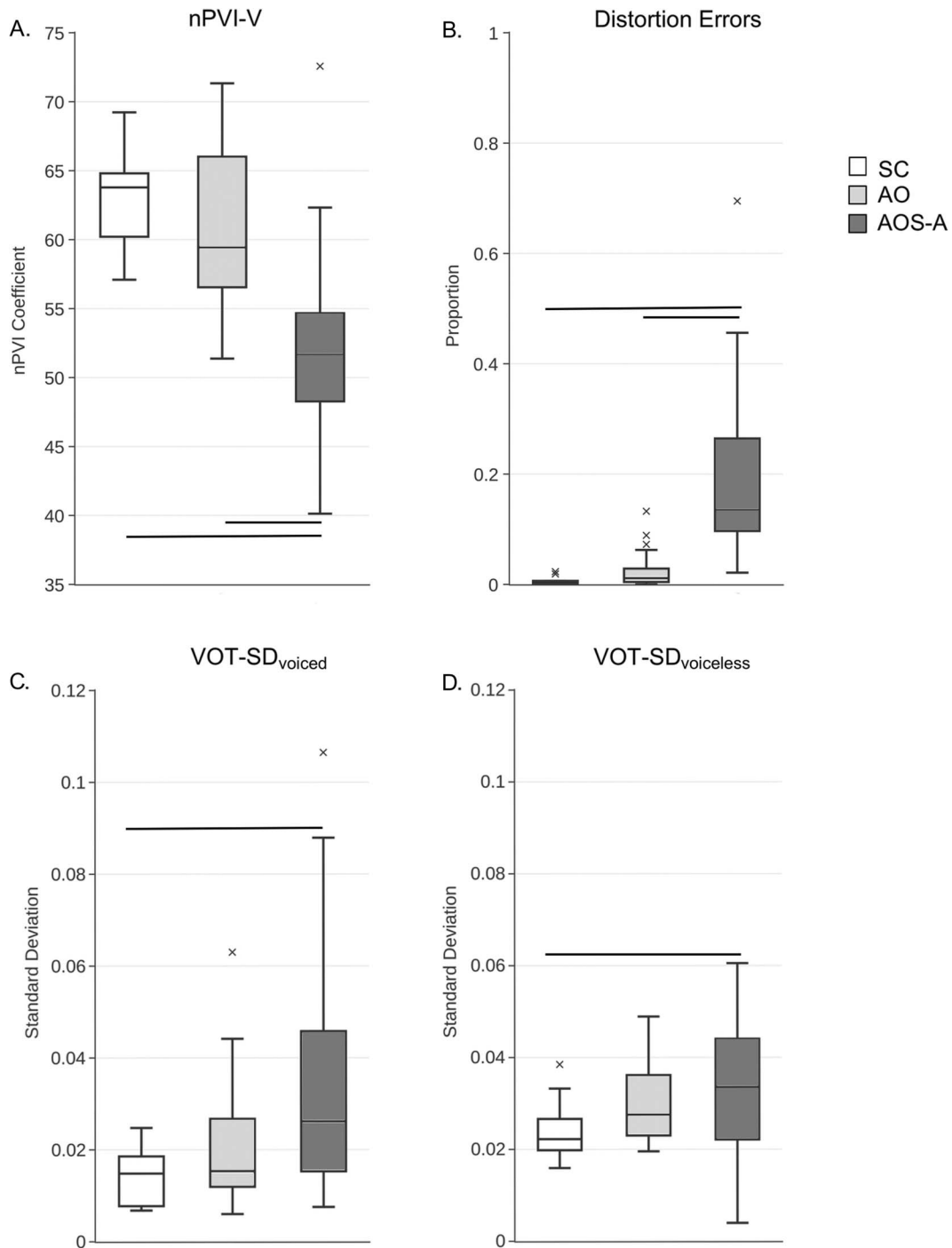
Only: 60.93 ± 5.51; AOS-Aphasia: 52.21 ± 7.11. A univariate ANOVA revealed a significant main effect of group, $F(2, 54) = 17.68$, $p < .001$, partial $\eta^2 = .40$. Bonferroni-corrected pairwise comparisons on the basis of the mean difference of nPVI-V scores show that the AOS-Aphasia group had significantly reduced nPVI-V scores when compared with the Stroke Control (mean difference = 10.34, $p < .001$) and Aphasia Only (mean difference = 8.72, $p < .001$) groups.

nPVI-V scores for the Stroke Control and Aphasia Only groups were not significantly different ($p = .95$). The nPVI-V values for the Stroke Control and Aphasia Only groups were in line with previously reported nPVI values for English speakers for connected speech samples (Arvaniti, 2012).

Distortion Errors

Group differences in the proportion of distortion errors were analyzed with a Kruskal–Wallis ANOVA, as the distributions of distortion errors violated the assumption of normality for all groups (Shapiro–Wilk $p < .05$ for all). Mean proportion of distortion errors ($\pm$ SDs) are as follows: Stroke Control: 0.005 ± 0.002; Aphasia Only: 0.025 ± 0.007; AOS-Aphasia: 0.188 ± 0.037. Box plots for each group are presented in Figure 2b. There was a main effect of group, $\chi^2(2) = 34.88$, $p < .0001$, with the Bonferroni-corrected Dunn's test showing that the AOS-Aphasia group had significantly more distortion errors than both the Stroke Control ($z = -32.26$, $p < .001$) and Aphasia Only ($z = -22.65$, $p < .001$) groups. There was no significant difference between the Stroke Control and Aphasia Only groups ($z = -9.64$, $p = .09$).

VOT Variability

The Aphasia Only and AOS-Aphasia groups' distributions for standard deviation of voice onset time (VOT-SD$_{voiced}$) did not meet the assumption of normality (Shapiro–Wilk $p < .05$), but all groups' distributions were normal for VOT-SD$_{voiceless}$. VOT data were analyzed separately—a Kruskal–Wallis one-way ANOVA was used to analyze the VOT-SD$_{voiced}$ variables, and a parametric one-way ANOVA was used for the VOT-SD$_{voiceless}$ comparisons.

Results of the Kruskal–Wallis one-way ANOVA of VOT-SD$_{voiced}$ showed a significant effect of group, $\chi^2(2) = 8.06$, $p = .02$. Pairwise comparisons with Mann–Whitney U tests revealed a significant difference between the AOS-Aphasia group when compared with the Stroke Control group (AOS-Aphasia mean rank = 19.75; Stroke Control mean rank = 11.08; $U = 55$, $z = -2.53$, $p < .01$), but not when compared with the Aphasia Only group (AOS-Aphasia mean rank = 25.83; Aphasia Only mean rank = 18.67.28; $U = 153$, $z = -1.86$, $p = .06$). There was no significant difference between VOT-SD$_{voiced}$ for the Stroke Control and Aphasia Only groups ($p = .11$).

Results of the one-way parametric ANOVA for VOT-SD$_{voiceless}$ were marginally significant, $F(2, 56) = 3.26$, $p = .049$, driven by a significant difference between the Stroke Control and AOS-Aphasia groups, $t(31) = 2.56$, $p = .016$. There were no other statistically significant comparisons. VOT variability for voiced and voiceless stop consonants is presented in Figures 2c and 2d, respectively.

EMS Variables

A one-way MANOVA was conducted to determine whether groups differed in amplitude energy for the three EMS modulation bands evaluated (1 Hz–2 Hz, 4 Hz–8 Hz, and 16 Hz–32 Hz). Results of the omnibus MANOVA showed that the three groups indeed differed significantly,

$F(6, 104) = 16.86$, $p < .001$, partial $\eta^2 = .49$. Follow-up univariate ANOVAs showed that these differences occurred in each of the three bands: 1 Hz–2 Hz: $F(2, 54) = 12.25$, $p < .001$, partial $\eta^2 = .31$; 4 Hz–8 Hz: $F(2, 54) = 6.56$, $p < .005$, partial $\eta^2 = .20$; 16 Hz–32 Hz: $F(2, 54) = 19.18$, $p < .001$, partial $\eta^2 = .42$. Post hoc tests revealed that the AOS-Aphasia group had greater amplitude energy in the 1 Hz–2 Hz and 16 Hz–32 Hz bands when compared with both the Stroke Control (mean difference = 0.06, $p < .001$) and Aphasia Only (mean difference = 0.04, $p < .005$) groups but less amplitude at the 4 Hz–8 Hz band when compared with the Stroke Control group (mean difference = -0.04, $p < .005$). The AOS-Aphasia and Aphasia Only groups did not differ significantly for the 4 Hz–8 Hz band ($p > .05$). There were no significant differences in amplitude modulation between the Stroke Control and Aphasia Only groups for the three bands evaluated ($p > .11$ for all comparisons). Figure 3 presents box plots for each modulation band by group.

Relationships With Overall AOS Severity

The relationship between ASRS AOS severity (quantified on a 0–4 scale, where a score of 0 means *AOS is absent* and a score of 4 means *AOS is marked in severity*; Strand et al., 2014) and each of the variables was inspected with Spearman correlation coefficients. All correlations, except for those with VOT-SD$_{voiced}$ and VOT-SD$_{voiceless}$, survived Bonferroni correction (p level = .05/7 comparisons = Bonferroni-corrected p level = .007). Results showed that overall AOS severity and the proportion of distortion errors were positively correlated ($r_s = .79$, $p < .001$), whereas nPVI-V was inversely related to AOS severity ($r_s = -.70$, $p < .001$). With respect to the EMS variables, more severe AOS is significantly correlated with greater amplitude energy in the 1 Hz–2 Hz ($r_s = .59$, $p < .001$) and 16 Hz–32 Hz ($r_s = .59$, $p < .001$) bands but less amplitude energy in the 4 Hz–8 Hz band ($r_s = -.46$, $p < .001$). Correlation coefficients between ASRS AOS severity and each variable are presented in Table 2.

Correlation With ASRS Items

Spearman correlation coefficients were also computed to relate each of the variables to scores for individual ASRS items. This was done for the AOS-Aphasia group only, as the other groups had too few items scored on the ASRS. The only correlations to reach statistical significance when Bonferroni corrected ($p < .0031$) were between (a) the 1 Hz–2 Hz band and ASRS Item 2.2 (*syllable segmentation across words in phrases/sentences*, $r_s = .64$, $p = .003$), (b) 4 Hz–8 Hz band and Item 2.4 (*slow overall speech rate*, $r_s = -.76$, $p < .0001$), (c) the proportion of distortion and Item 2.1 (*syllable segmentation within words > 1 syllable*, $r_s = .65$, $p = .003$), and (d) the proportion of distortion errors and Item 2.2 (*syllable segmentation across words in phrases/sentences*, $r_s = .68$, $p = .001$). There was a significant correlation between nPVI-V and Item 1.4 (*increased sound distortions…with increased rate*), but this did not remain significant when subject to Bonferroni correction ($r_s = -.48$,

Figure 3. Amplitude energy for each of the EMS bands tested. Horizontal lines indicate significant group differences at the Bonferroni-corrected *p* level of .0167. EMS = envelope modulation spectrum; SC = Stroke Control; AO = Aphasia Only; AOS-A = Apraxia of Speech with Concomitant Aphasia.
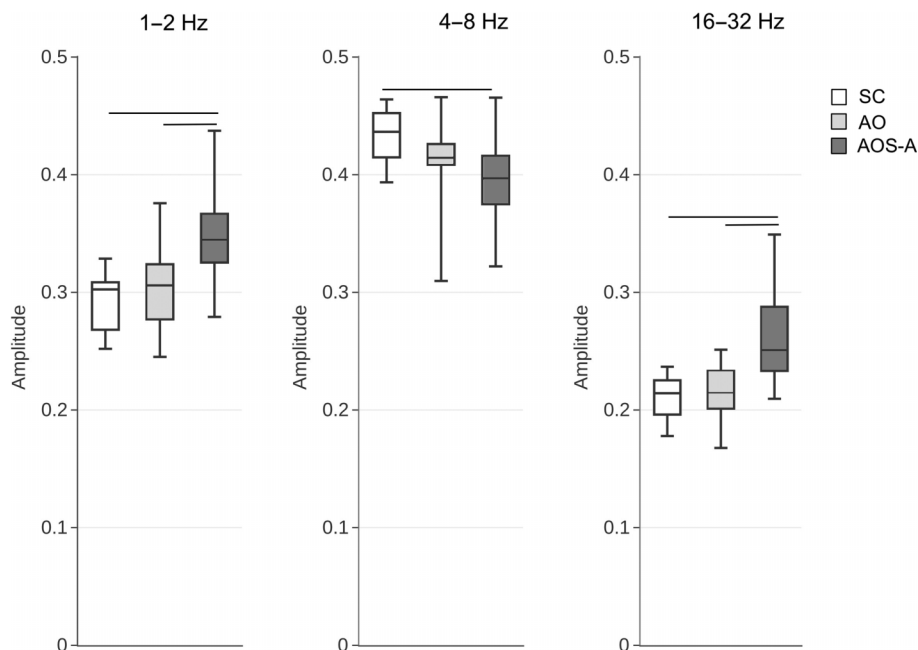
p = .04). There were no significant correlations between ASRS Items and VOT variability or the 16 Hz–32 Hz band. Table S1 (Supplemental Material S5) presents all correlation coefficients between predictor variables and ASRS item scores.

Interim Summary

Results from the above analyses show that the AOS-Aphasia group differed significantly from (a) the Stroke Control group on all measures and (b) the Aphasia Only group on all measures except for amplitude energy in the 4 Hz–8 Hz modulation band and VOT variability. The

Table 2. Spearman correlation coefficients between all variables and ASRS AOS severity scores (quantified on a 0–4 scale).

Predictor	Correlation coefficient
Proportion distortions	$r_s = .79$, $p < .001$*
nPVI-V	$r_s = -.70$, $p < .001$*
VOT-SD$_{voiced}$	$r_s = .35$, $p = .01$
VOT-SD$_{voiceless}$	$r_s = -.28$, $p = .04$
1 Hz–2 Hz band	$r_s = .59$, $p < .001$*
4 Hz–8 Hz band	$r_s = -.46$, $p < .001$*
16 Hz–32 Hz band	$r_s = .59$, $p < .001$*

Note. ASRS = Apraxia of Speech Rating Scale; AOS = Apraxia of Speech; nPVI-V = normalized Pairwise Variability Index–Vowels; VOT-SD = standard deviation of voice onset time.

*Correlations significant at the Bonferroni-corrected $p = .007$ ($p = .05/7$ comparisons).

Aphasia Only and Stroke Control groups did not significantly differ on any measures investigated.
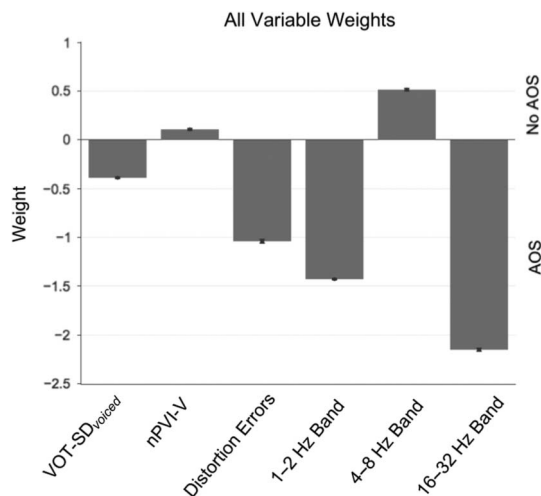
Classification Analysis

Discriminant analyses were completed by comparing the individuals with AOS (*n* = 20) to (a) all participants without AOS (*n* = 37; collapsed into one group) and (b) the Aphasia Only group (*n* = 24). Models were run first with all variables, and then a series of LDA models were run with each variable, independently, to determine the amount of variance accounted for by each variable on its own. Results of each model are detailed in the following sections. Note that all percentages reported for LDA accuracy reflect accuracy following the leave-one-out cross-validation procedure. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) are reported using the ASRS guidelines as a reference (i.e., total score of ≥ 8, at least one item scored from Section 1.1–1.6; Strand et al., 2014).

All Variables: AOS-Aphasia Versus All Others

The model that included all variables was 96.5% accurate in diagnostic classification ($p = 6.3 \times 10^{-10}$; sensitivity = 100%; specificity = 94.6%; PPV = 90.9%; NPV = 100%). As presented in Figure 4, higher amplitude energy in the 1 Hz–2 Hz and 16 Hz–32 Hz bands was associated with AOS group membership, but higher nPVI-V coefficients and 4 Hz–8 Hz amplitude energies were associated with the absence of AOS.

Figure 4. Results from the linear discriminant analysis (LDA) with all variables and all participants included. Variables with negative weights indicate that a higher score on that variable is predictive of AOS diagnosis (according to the Apraxia of Speech Rating Scale [ASRS]). Variables with positive weights indicate that a higher score on that variable was not associated with AOS diagnosis. AOS = apraxia of speech; nPVI-V = normalized Pairwise Variability Index–Vowels; VOT-SD = standard deviation of voice onset time.

Table 3. Correlation coefficients (r denotes Pearson; r_s denotes Spearman) between LDA input variables and WAB AQ scores.

Predictor	Correlation coefficient
1 Hz–2 Hz band	$r = -.20, p = .14$
4 Hz–8 Hz band	$r = .34,^* p = .009$
16 Hz–32 Hz band	$r = -.47,^* p < .001$
VOT-SD$_{voiced}$	$r_s = -.09, p = .51$
VOT-SD$_{Voiceless}$	$r = -.23, p = .08$
nPVI-V	$r = .22, p = .09$
Distortion errors	$r_s = -.48,^* p < .001$

Note. LDA = linear discriminant analysis; WAB AQ = Western Aphasia Battery Aphasia Quotient; VOT-SD = standard deviation of voice onset time; nPVI-V = normalized Pairwise Variability Index–Vowels.

*Correlations significant at the Bonferroni-corrected p value of $p = .007$ ($p = .05/7$ comparisons).

and Aphasia Only groups. Results are presented in Tables 4 and 5.

All Variables: AOS-Aphasia Versus Aphasia Only

LDA classification accuracy was 95.45% ($p = 9.85 \times 10^{-11}$; sensitivity = 100%; specificity = 91.7%; PPV = 90.9%; NPV = 100%) when only the AOS-Aphasia and Aphasia Only groups were compared. The relative weight and direction of each variable did not differ from the (above) model where all participants were included (see Figure 4).

The AOS-Aphasia group had significantly greater aphasia severity than the other two groups, but inclusion of WAB AQ in all of the aforementioned models did not change prediction accuracy, and the WAB AQ had the least weight in discrimination (< -0.28 across models). Some of the variables included were correlated with WAB AQ scores (all correlation coefficients $| r | < .48$, as presented in Table 3), but the finding that inclusion of WAB AQ did not improve diagnostic classification in the LDA suggests that WAB AQ does not contribute any additional predictive information that has not already been captured by other variables in our model (see Figure S3 in Supplemental Material S6).

Individual Variables LDA

When each variable was inspected individually, all variables, except for VOT variability, yielded a significant prediction model ($p < .05$). The distortion errors and 16 Hz–32 Hz variables each yielded the highest classification accuracies. On its own, nPVI-V had the highest sensitivity and NPV, whereas distortion errors had the highest specificity and PPV. Results were similar when all groups were combined, or when restricting analyses to the AOS-Aphasia

Discussion

The purpose of this study was to determine which acoustic speech measures aid in the diagnosis of AOS and to identify the measures that account for the greatest variance in diagnostic classification. Results showed that the objective measures included here indeed distinguish individuals with AOS and concomitant aphasia from those with aphasia only. Each of the individual objective measures (except for VOT variability) had moderate to high sensitivity and specificity for diagnostic classification when confirmed with a clinical diagnosis obtained from the ASRS (see Tables 4 and 5).

Unlike prior work (e.g., Ballard et al., 2016; Vergis et al., 2014), we obtained each production measure from connected speech samples rather than isolated word/sentence productions that were constant across speakers. The high classification accuracy, along with use of relatively unconstrained materials, suggests that these measures are capturing important production features that occur in longer connected speech samples. This high accuracy is also an indication of how robust these measures are against variability in spoken speech.

Recent studies aimed at improving differential diagnosis through objective measures have revealed that the PVI coefficient may predict AOS diagnosis with high sensitivity and specificity. Ballard et al. (2016) demonstrated that PVI coefficients from words with a weak–strong stress pattern and scores from the "increasing word-length" subtest of the ABA-2 (evaluates production upon trials where words increase in length/complexity, i.e., *cat, catapult, catastrophe;* see ABA-2 for additional details and word list) reliably differentiated speakers with AOS and aphasia from those with aphasia only. PVI coefficients used in this study were derived from connected speech samples, whereas Ballard et al. (2016) and Vergis et al. (2014) computed PVI

Table 4. Classification accuracy, sensitivity, specificity, and positive and negative predictive values (PPV, NPV, respectively) of each individual variable for the AOS-Aphasia group relative to all other participants.

Variable	Accuracy	Sensitivity	Specificity	PPV	NPV	Criterion score
VOT-SD$_{voiced}$	68 (n.s.)	50	78	55	74	0.03
nPVI-V	79	**85**	76	65	**90**	57
Distortion errors	**84**	60	**97**	**93**	82	11%
1 Hz–2 Hz	81	**80**	81	70	**88**	0.32
4 Hz–8 Hz	77	65	84	68	82	0.65
16 Hz–32 Hz	**82**	70	**89**	**78**	85	0.24

Note. All variables yielded significant predictive models with the exception of the VOT-SD$_{voiced}$ variable (n.s. = nonsignificant). The last column reports the criterion score for the given variable's sensitivity/specificity. The two predictors with highest values for each metric are presented in bold and underlined text. AOS = apraxia of speech; VOT-SD = standard deviation of voice onset time; nPVI-V = normalized Pairwise Variability Index–Vowels.

from single-word productions, perhaps explaining differences in mean PVI values for participant groups in this study. Ballard et al.'s (2016) study also differed from the current study in other key aspects: (a) A different method of prediction was used; (b) Their statistical model included a large number of variables in addition to measures of speech production (i.e., demographics, comprehension, naming, working memory, reading performance, auditory word discrimination, and oral motor movement abilities). Despite these differences, the results from the current study corroborate Ballard et al.'s findings with an independent and comparably large sample—confirming that certain features are indeed different in individuals with and without AOS and can be used to discriminate between groups reliably. We should also emphasize that our high classification accuracy (> 90% for the full LDA model) was obtained in a leave-one-out fashion, where the "to-be diagnosed" individual was excluded from the model estimation procedure. This cross-validation procedure reflects how well our classification model generalizes to new data, providing a measure of confidence that these results are not attributed to over-fitting the statistical model.

The inclusion of the EMS variables in this study expands upon efforts to identify other objective behavioral measures to differentiate individuals with AOS and concomitant aphasia from speakers with only aphasia or no

impairment (as detected by the WAB or ASRS). To our knowledge, this is the first study that has investigated amplitude modulation spectra in the classification of AOS. The following sections will include a brief discussion of the clinical and theoretical relevance of the EMS measures, how the EMS measures relate to the other measures included in this study, and suggestions for future research.

Relevance of EMS Measures in AOS Evaluation

It should be emphasized that the EMS measures are associated with how speech is produced, but not necessarily with one specific aspect of speech. As discussed in the Introduction and Method sections, slower rate modulations (1 Hz–2 Hz) reflect speech prosody, the 4 Hz–8 Hz range is associated with syllabic rate, and the higher frequency band (16 Hz–32 Hz) corresponds to consonant features. As shown in Figure 3, the AOS-Aphasia group had significantly greater amplitude energy in the 1 Hz–2 Hz and 16 Hz–32 Hz bands when compared to the Stroke Control and Aphasia Only groups (who did not differ significantly from each other). This finding suggests that speakers with AOS place equal emphasis on words within phrases and sentences (i.e., greater 1 Hz–2 Hz amplitudes) and emphasizes individual sounds within each word (i.e., greater 16 Hz–32 Hz amplitudes). The former likely reflects reduced

Table 5. Classification accuracy, sensitivity, specificity, and positive and negative predictive values (PPV, NPV, respectively) of each individual variable for the AOS-Aphasia group relative to the Aphasia Only group.

Variable	Accuracy	Sensitivity	Specificity	PPV	NPV	Criterion score
VOT-SD$_{voiced}$	63 (n.s.)	50	75	63	64	0.03
nPVI-V	75	**85**	67	68	**84**	56.8
Distortion errors	**80**	60	**96**	**92**	74	11%
1 Hz–2 Hz	75	**70**	79	74	76	0.33
4 Hz–8 Hz	71	60	79	71	70	0.40
16 Hz–32 Hz	**80**	**70**	**88**	**82**	**78**	0.24

Note. All variables yielded significant predictive models with the exception of the VOT-SD$_{voiced}$ variable (n.s. = nonsignificant). The last column reports the criterion score for the given variable's sensitivity/specificity. The two predictors with highest values for each metric are presented in bold and underlined text. AOS = apraxia of speech; VOT-SD = standard deviation of voice onset time; nPVI-V = normalized Pairwise Variability Index–Vowels.

variance in prosodic patterns, whereas the latter may reflect articulatory problems and stalling on individual speech sounds, that is, less smooth transitions between articulations.

Inspection of correlation coefficients (Figure S2) shows that the 1 Hz–2 Hz band was inversely correlated with nPVI. It is important to note that correlations with each ASRS item show that, for the participants with AOS, amplitude energy in the 1 Hz–2 Hz band was correlated with ASRS items that reflect perceptual evaluation of prosody (e.g., syllable segmentation across words in phrases/sentences, Item 2.2), but nPVI-V was not. It has been argued that segmental duration measures, like nPVI, ignore important information about the acoustic signal and may not relate to perceptual evaluation of prosody (Lowit, 2014; Tilsen & Johnson, 2008). One clinical and theoretical implication is that measures derived from the speech signal, such as EMS measures, may be more informative in speech analysis (Lowit, 2014; Tilsen & Johnson, 2008).

With regard to phonetic features, the correlation coefficients restricted to the AOS-Aphasia group (Figure S2) show that the 16 Hz–32 Hz band is correlated with distortion errors ($r = -.47$, $p < .05$). This suggests that greater emphasis on consonant features during connected speech results in fewer distortion errors. Further research should examine this relationship as it pertains to different types of sound level errors. Regarding the 1 Hz–2 Hz band, it has been speculated that dysprosody may be a compensatory mechanism for articulatory impairments (e.g., Darley et al., 1975). However, the significant (positive) correlation between distortions and 1 Hz–2 Hz ($r = .51$, $p < .05$) for the AOS participants suggests that greater emphasis on prosodic features does not necessarily attenuate the frequency of distortion errors. Rather, this relationship between prosody (1 Hz–2 Hz band) and distortion errors may be driven by overall AOS severity (see Table 2), where dysprosody is a primary deficit that results from impaired motor planning/programming (e.g., Rogers, 1997).

Refining Objective Measures for AOS Diagnosis: Future Directions

The diagnostic value of these measures remains in the early stages of investigation, but the results obtained here are promising. When all variables were included in the LDA, the EMS variables were by far the predictors with the greatest weight in clinical classification, even though the univariate ANOVAs for each of the three EMS measures had smaller effect sizes and mean group differences relative to nPVI-V and the proportion of distortion errors (see also Figures 2a–2d). This apparent paradox can be explained by the fact that the LDA is a multivariate analysis that simultaneously considers (a) correlations between measures and (b) mean differences between groups (i.e., "center distances"), meaning a given variable can have a higher weight in classification even if there is a substantial overlap in that variable's value across the classes, but at the same time, the correlations between this variable with the other variables

can be used to inform the separation function (see Haynes & Rees, 2006 for a detailed discussion). Because LDA predicts the group membership from a combination of variables, the weight of a variable within this combination reflects its importance when it is combined with other variables. As our results suggest, more than one measure should be obtained for the classification of AOS. As evident in Tables 4–5, several measures have relatively high accuracy in prediction when analyzed individually, but when inspecting results from the LDA model with all variables, the combination of variables yielded higher predictive value, with 1 Hz–2 Hz (a measure of prosody) and 16 Hz–32 Hz (a measure of consonant features) being the strongest predictors. This is rather similar to Ballard et al.'s finding (2016) that measures of prosody (nPVI-V), along with a measure of sound errors, were highly predictive of AOS. The benefit to further exploration of the EMS measures is that these two top predictors can be obtained with an automated procedure. Further work should investigate the most efficient combination of variables for prediction.

Based on these findings, future research investigating objective classification of AOS should consider implementing amplitude modulation measures, as replication of these results across clinical and research sites could have important, and practical, clinical applications. First, these variables had high accuracy in AOS classification, and variables representing low amplitude modulation rates have also demonstrated similar accuracy in discriminating between the dysarthria subtypes (Liss et al., 2010). Second, we suggest that the EMS measures could be relatively easily automated for widespread clinical use. We acknowledge that the speech samples that were analyzed in this study required some preprocessing (i.e., removing noise unrelated to the participant), but this process may be avoided if speech samples are obtained in a quiet environment and experimenters/clinicians refrain from providing verbal cues/prompts. MATLAB scripts used for the EMS analyses may be packaged in formats more accessible to clinicians (i.e., via downloadable apps). Future research should also consider the influence of different analysis settings (e.g., window durations) or the type of elicitation materials on classification outcomes. In light of our present findings, we will continue to work toward the development of such applications.

Limitations

The lack of a reliable gold standard definition or tool for diagnosing AOS has remained one of the biggest challenges to its clinical management. Moreover, inherent in any study where a perceptual evaluation is used as the gold standard, the issues of circularity and nonindependence between perceptual ratings and the measures under investigation are of concern. The present study diagnosed AOS on the basis of the ASRS, a scale that has been shown to have high validity when compared with expert diagnosis (Strand et al., 2014). Nevertheless, the ASRS is inherently a perceptual scale and has not yet been validated in a multisite

study. Because ASRS diagnoses were used to confirm the classification accuracy of the LDA model, overall accuracy of the LDA could have been influenced by ASRS AOS diagnosis itself. To address this concern post hoc, we applied an additional unsupervised approach that does not use clinical classification as a guide (k-means cluster approach, where k = 2 "forced" groups) to evaluate further our original LDA model. We found that the predictor variables classified participants into two groups. When comparing ASRS classification to these groups, this unsupervised approach grouped 16 of the 20 with AOS into the same group, for an overall accuracy of 88%. In the absence of an uncontroversial gold standard, this may either mean that (a) certain characteristics of AOS are not captured by our acoustic measures, but only by the perceptual ratings on the ASRS, or that (b) the diagnosis of AOS on the basis of the perceptual ASRS ratings is erroneous in at least 12% of cases. At present, there is no means of establishing which is true.

Aside from the challenges imposed by studying poststroke AOS in the context of aphasia (e.g., Graff-Radford et al., 2014), another challenge to the differential diagnosis of AOS is the presence of dysarthria. As in AOS, speakers with dysarthria may demonstrate atypical prosody and speech sound distortions (Strand et al., 2014). However, unless basal ganglia and/or brain stem structures are affected, dysarthria often resolves after a unilateral cortical stroke (Duffy, 2005). In cases of unilateral upper motor neuron dysarthria that persists, effects tend to be mild (Duffy, 2005). This can be exemplified by the rate/severity of dysarthria in the current sample (n = 3 individuals with a diagnosis of dysarthria in the Aphasia Only group, n = 6 in the AOS-Aphasia group; mean ASRS dysarthria severity score < 1 for the Aphasia Only and AOS-Aphasia groups). Nevertheless, the validity of measures intended to quantify speech production deficits that occur due to a neurologic injury should also account for the possibility of concomitant dysarthria. Because few participants in this sample presented with dysarthria, we were unable to determine the extent to which the measures used here could also distinguish those with dysarthria from the speakers with AOS. However, there is precedent for using a more fine-grained EMS analysis for the classification of dysarthria subtypes (Liss et al., 2010), indicating the potential for differential diagnosis between AOS and dysarthria using similar acoustic metrics.

Similarly, further research is necessary regarding how these measures fare in more "difficult" diagnostic decision-making cases. For example, conduction aphasia may often be misdiagnosed as AOS due to frequent phonemic errors and frequent attempts at self-correction (Haley et al., 2013). In addition, both AOS and Broca's aphasia result from similar patterns of brain damage (e.g., Hillis et al., 2004; Richardson, Fillmore, Rorden, LaPointe, & Fridriksson, 2012) and are characterized by similar sounding impairments, such that individuals with Broca's aphasia are often falsely classified with AOS due to the effortful struggle with speech and reduced fluency that characterize both impairments. The current study included a sample size of N = 57, yielding sufficient power for meaningful classification accuracy.

However, as evidenced by Table 1, direct comparisons between subgroups of different types of aphasia would result in analyses with small sample sizes. Further research should consider these more challenging clinical questions.

Conclusions

With over 90% classification accuracy, the acoustic measures of this study are promising and warrant further analysis and replication for the development of these measures as behavioral markers of AOS. Pending further research, the measures investigated here may provide an objective way to estimate prognosis and monitor patients' responses to treatment throughout the course of speech/language therapy, as severity/intelligibility judgments can be biased due to clinician familiarity with a patient (Kent, 1996).

References

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics, 40*(3), 351–373.

Ballard, K. J., Azizi, L., Duffy, J., McNeil, M. R., Halaki, M., O'Dwyer, N., . . . Robin, D. A. (2016). A predictive model for diagnosing stroke-related apraxia of speech. *Neuropsychologia, 81,* 129–139.

Ballard, K. J., Granier, J. P., & Robin, D. A. (2000). Understanding the nature of apraxia of speech: Theory, analysis and treatment. *Aphasiology, 14*(10), 969–995.

Basilakos, A., Rorden, C., Bonilha, L., Moser, D., & Fridriksson, J. (2015). Patterns of poststroke brain damage that predict speech production errors in apraxia of speech and aphasia dissociate. *Stroke, 46*(6), 1561–1566. https://doi.org/10.1161/STROKEAHA.115.009211

Boersma, P., & Weenink, D. (2001). Praat: A system for doing phonetics by computer [Computer program]. Retrieved from http://www.praat.org

Buckingham, H. W., & Yule, G. (1987). Phonemic false evaluation: Theoretical and clinical aspects. *Clinical Linguistics & Phonetics, 1*(2), 113–125.

Canter, G. J., Trost, J. E., & Burns, M. S. (1985). Contrasting speech patterns in apraxia of speech and phonemic paraphasia. *Brain and Language, 24*(2), 204–222.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284–290.

Code, C. (1998). Models, theories and heuristics in apraxia of speech. *Clinical Linguistics & Phonetics, 12*(1), 47–65.

Collins, M., Rosenbek, J. C., & Wertz, R. T. (1983). Spectrographic analysis of vowel and word duration in apraxia of speech. *Journal of Speech, Language, and Hearing Research, 26*(2), 224–230.

Crouzet, O., & Ainsworth, W. A. (2001, September). *On the various influences of envelope information on the perception of speech in adverse conditions: An analysis of between-channel*

envelope correlation. Paper presented at the Workshop on Consistent and Reliable Acoustic Cues for Sound Analysis, Aalborg, Denmark.

Cunningham, K. T., Haley, K. L., & Jacks, A. (2016). Speech sound distortions in aphasia and apraxia of speech: Reliability and diagnostic significance. *Aphasiology, 304*(4), 396–413.

Dabul, B.L. (1979). *Apraxia Battery for Adults*. Tigard, OR: C.C. Publications.

Dabul, B. L. (2000). *Apraxia Battery for Adults–Second Edition*. Dallas, TX: Pro-Ed.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1975). *Motor speech disorders*. Philadelphia, PA: W.B. Saunders Co.

Duffy, J. (2005). *Motor speech disorders: Substrates, differential diagnosis and management*. St. Louis, MO: Mosby.

Dunn, O. J. (1964). Multiple comparisons using rank sums. *Technometrics, 6*(3), 241–252.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics, 7*(2), 179–188.

Fromm, D., Abbs, J. H., McNeil, M. R., & Rosenbek, J. C. (1982). Simultaneous perceptual-physiological method for studying apraxia of speech. In R. H. Brookshire, (Ed.), *Clinical aphasiology conference proceedings* (pp. 65–80). Minneapolis, MN: BRK Publishers.

Galluzzi, C., Bureca, I., Guariglia, C., & Romani, C. (2015). Phonological simplifications, apraxia of speech and the interaction between phonological and phonetic processing. *Neuropsychologia, 71*, 64–83.

Ghitza, O. (2011). Linking speech perception and neurophysiology: Speech decoding guided by cascaded oscillators locked to the input rhythm. *Frontiers in Psychology, 2*, 130.

Ghitza, O., & Greenberg, S. (2009). On the possible role of brain rhythms in speech perception: Intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica, 66*(1–2), 113–126.

Giraud, A. L., Kleinschmidt, A., Poeppel, D., Lund, T. E., Frackowiak, R. S., & Laufs, H. (2007). Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron, 56*(6), 1127–1134.

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience, 15*(4), 511–517.

Goodglass, H., Kaplan, E., & Barresi, B. (2000). *Boston Diagnostic Aphasia Examination–Third Edition (BDAE-3)*. San Antonio, TX: The Psychological Corporation.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. *Papers in Laboratory Phonology, 7*(515–546).

Graff-Radford, J., Jones, D. T., Strand, E. A., Rabinstein, A. A., Duffy, J. R., & Josephs, K. A. (2014). The neuroanatomy of pure apraxia of speech in stroke. *Brain and Language, 129*, 43–46. https://doi.org/10.1016/j.bandl.2014.01.004

Haley, K. L., Jacks, A., & Cunningham, K. T. (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. https://doi.org/10.1044/1092-4388(2012/12-0161)

Haley, K. L., Jacks, A., de Riesthal, M., Abou-Khalil, R., & Roth, H. L. (2012). Toward a quantitative basis for assessment and diagnosis of apraxia of speech. *Journal of Speech, Language, and Hearing Research, 55*(5), S1502–S1517. https://doi.org/10.1044/1092-4388(2012/11-0318)

Hall, J. W., & Grose, J. H. (1993). Monaural envelope correlation perception in listeners with normal hearing and cochlear impairment. *Journal of Speech, Language, and Hearing Research, 36*(6), 1306–1314.

Haynes, J. D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience, 7*(7), 523–534.

Hillis, A. E., Work, M., Barker, P. B., Jacobs, M. A., Breese, E. L., & Maurer, K. (2004). Re-examining the brain regions crucial for orchestrating speech articulation. *Brain, 127*(7), 1479–1487.

Itoh, M., & Sasanuma, S. (1984). Articulatory movements in apraxia of speech. In J. C. Rosenbek, M. R. McNeil, & A. E. Aronson (Eds.), *Apraxia of speech: Physiology, acoustics, linguistics, management* (pp. 135–165). San Diego, CA: College-Hill Press.

Itoh, M., Sasanuma, S., Tatsumi, I. F., Murakami, S., Fukusako, Y., & Suzuki, T. (1982). Voice onset time characteristics in apraxia of speech. *Brain and Language, 17*(2), 193–210.

Itoh, M., Sasanuma, S., & Ushijima, T. (1979). Velar movements during speech in a patient with apraxia of speech. *Brain and Language, 7*(2), 227–239.

Jacks, A., Mathes, K. A., & Marquardt, T. P. (2010). Vowel acoustics in adults with apraxia of speech. *Journal of Speech, Language, and Hearing Research, 53*(1), 61–74.

Kent, R. D. (1996). Hearing and believing some limits to the auditory-perceptual assessment of speech and voice disorders. *American Journal of Speech-Language Pathology, 5*(3), 7–23.

Kent, R. D., & Rosenbek, J. C. (1983). Acoustic patterns of apraxia of speech. *Journal of Speech, Language, and Hearing Research, 26*(2), 231–249.

Kertesz, A. (2006). *Western Aphasia Battery–Revised*. San Antonio, TX: Pearson.

Leong, V., Stone, M. A., Turner, R. E., & Goswami, U. (2014). A role for amplitude modulation phase relationships in speech rhythm perception. *The Journal of the Acoustical Society of America, 136*(1), 366–381.

Liss, J. M., LeGendre, S., & Lotto, A. J. (2010). Discriminating dysarthria type from envelope modulation spectra. *Journal of Speech, Language, and Hearing Research, 53*(5), 1246–1255.

Liss, J. M., White, L., Mattys, S. L., Lansford, K., Lotto, A. J., Spitzer, S. M., & Caviness, J. N. (2009). Quantifying speech rhythm abnormalities in the dysarthrias. *Journal of Speech, Language, and Hearing Research, 52*(5), 1334–1352.

Lowit, A. (2014). Quantification of rhythm problems in disordered speech: A re-evaluation. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 369*(1658), 20130404.

MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences, 21*(04), 499–511.

McNeil, M. R., & Adams, S. (1991). A comparison of speech kinematics among apraxic, conduction aphasic, ataxic dysarthria, and normal geriatric speakers. *Clinical Aphasiology, 19*, 279–294.

McNeil, M. R., Liss, J. M., Tseng, C.-H., & Kent, R. D. (1990). Effects of speech rate on the absolute and relative timing of apraxic and conduction aphasic sentence production. *Brain and Language, 38*(1), 135–158.

McNeil, M. R., Pratt, S. R., & Fossett, T. R. D. (2004). The differential diagnosis of apraxia of speech. In B. Maassen, R. Kent, H. Peters, P. van Lieshout, & W. Hulstijn (Eds.), *Speech motor control in normal and disordered speech* (pp. 389–413). Oxford, United Kingdom: Oxford University Press.

McNeil, M. R., Robin, D. A., & Schmidt, R. A. (1997). Apraxia of speech: Definition, differentiation, and treatment. In M. R. McNeil (Ed.), *Clinical management of sensorimotor speech disorders* (pp. 311–344). New York, NY: Thieme.

Miller, N. (1995). Pronunciation errors in acquired speech disorders: The errors of our ways. *International Journal of Language and Communication Disorders, 30*(3), 346–361.

Moser, D., Basilakos, A., Fillmore, P., & Fridriksson, J. (2016). Brain damage associated with apraxia of speech: Evidence from case studies. *Neurocase, 22*, 1–11. https://doi.org/10.1080/13554794.2016.1172645

Mumby, K., Bowen, A., & Hesketh, A. (2007). Apraxia of speech: How reliable are speech and language therapists' diagnoses? *Clinical Rehabilitation, 21*(8), 760–767.

Murray, L. L., Holland, A. L., & Beeson, P. M. (1998). Spoken language of individuals with mild fluent aphasia under focused and divided-attention conditions. *Journal of Speech, Language, and Hearing Research, 41*(1), 213–227.

Odell, K., McNeil, M. R., Rosenbek, J. C., & Hunter, L. (1990). Perceptual characteristics of consonant production by apraxic speakers. *Journal of Speech and Hearing Disorders, 55*(2), 345–359.

Odell, K., McNeil, M. R., Rosenbek, J. C., & Hunter, L. (1991). Perceptual characteristics of vowel and prosody production in apraxic, aphasic, and dysarthric speakers. *Journal of Speech, Language, and Hearing Research, 34*(1), 67–80.

Ogar, J., Willock, S., Baldo, J., Wilkins, D., Ludy, C., & Dronkers, N. (2006). Clinical and anatomical correlates of apraxia of speech. *Brain and Language, 97*(3), 343–350. https://doi.org/10.1016/j.bandl.2006.01.008

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America, 32*(6), 693–703.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time.' *Speech Communication, 41*(1), 245–255.

Richardson, J. D., Fillmore, P., Rorden, C., LaPointe, L. L., & Fridriksson, J. (2012). Re-establishing Broca's initial findings. *Brain and Language, 123*(2), 125–130.

Robin, D. A., Bean, C., & Folkins, J. W. (1989). Lip movement in apraxia of speech. *Journal of Speech, Language, and Hearing Research, 32*(3), 512–523.

Rogers, M. A. (1997). The vowel lengthening exaggeration effect in speakers with apraxia of speech: Compensation, artifact, or primary deficit? *Aphasiology, 11*(4–5), 433–445.

Rosen, S. (1992). Temporal information in speech: Acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 336*(1278), 367–373.

Rosenbek, J., Kent, R., & Lapointe, L. (1984). Apraxia of speech: An overview and some perspectives. In J. C. Rosenbek, M. R. McNeil, & A. E. Aronson (Eds.), *Apraxia of speech: Physiology, acoustics, linguistics, management* (pp. 1–72). San Diego, CA: College-Hill Press.

Seddoh, S. A., Robin, D. A., Sim, H. S., Hageman, C., Moon, J. B., & Folkins, J. W. (1996). Speech timing in apraxia of speech versus conduction aphasia. *Journal of Speech and Hearing Research, 39*(3), 590–603.

Shankweiler, D., & Harris, K. S. (1966). An experimental approach to the problem of articulation in aphasia. *Cortex, 2*(3), 277–292.

Square-Storer, P., & Appledoorn, S. (1991). An acoustic study of apraxia of speech in patients with different lesion loci. In C. Moore, K. Yorkston, & D. Beukelman (Eds.), *Dysarthria and apraxia of speech: Perspectives on management* (pp. 271–288). Baltimore, MD: Paul Brookes Publishing.

Strand, E. A., Duffy, J. R., Clark, H. M., & Josephs, K. (2014). The apraxia of speech rating scale: A tool for diagnosis and description of apraxia of speech. *Journal of Communication Disorders, 51*, 43–50. https://doi.org/10.1016/j.jcomdis.2014.06.008

Thomas, E. R., & Carter, P. M. (2006). Prosodic rhythm and African American English. *English World-Wide, 27*(3), 331–355.

Tilsen, S., & Johnson, K. (2008). Low-frequency Fourier analysis of speech rhythm. *The Journal of the Acoustical Society of America, 124*(2), EL34–EL39.

Varley, R. A., & Whiteside, S. P. (2001). What is the underlying impairment in acquired apraxia of speech. *Aphasiology, 15*(1), 39–49.

Vergis, M. K., Ballard, K. J., Duffy, J. R., McNeil, M. R., 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.

Wambaugh, J. L., Wright, S., Nessler, C., & Mauszycki, S. C. (2014). Combined aphasia and apraxia of speech treatment (CAAST): Effects of a novel therapy. *Journal of Speech, Language, and Hearing Research, 57*(6), 2191–2207.

Wertz, R., LaPointe, L., & Rosenbek, J. (1984). *Apraxia of speech: The disorders and its management*. New York, NY: Grune & Stratton.

Ziegler, W. (2005). A nonlinear model of word length effects in apraxia of speech. *Cognitive Neuropsychology, 22*(5), 603–623.

Ziegler, W. (2009). Modeling the architecture of phonetic plans: Evidence from apraxia of speech. *Language and Cognitive Processes, 24*, 631–661.

Ziegler, W., & von Cramon, D. (1985). Anticipatory coarticulation in a patient with apraxia of speech. *Brain and Language, 26*(1), 117–130.

Ziegler, W., & von Cramon, D. (1986a). Disturbed coarticulation in apraxia of speech: Acoustic evidence. *Brain and Language, 29*(1), 34–47.

Ziegler, W., & von Cramon, D. (1986b). Timing deficits in apraxia of speech. *European Archives of Psychiatry and Neurological Sciences, 236*(1), 44–49.