Discriminating Dysarthria Type From Envelope Modulation Spectra

RESEARCH NOTE

Julie M. Liss
Arizona State University, Tempe

Sue LeGendre
Andrew J. Lotto
University of Arizona, Tucson

Purpose: Previous research demonstrated the ability of temporally based rhythm metrics to distinguish among dysarthrias with different prosodic deficit profiles (J. M. Liss et al., 2009). The authors examined whether comparable results could be obtained by an automated analysis of speech envelope modulation spectra (EMS), which quantifies the rhythmicity of speech within specified frequency bands.
Method: EMS was conducted on sentences produced by 43 speakers with 1 of 4 types of dysarthria and healthy controls. The EMS consisted of the spectra of the slow-rate (up to 10 Hz) amplitude modulations of the full signal and 7 octave bands ranging in center frequency from 125 to 8000 Hz. Six variables were calculated for each band relating to peak frequency and amplitude and relative energy above, below, and in the region of 4 Hz. Discriminant function analyses (DFA) determined which sets of predictor variables best discriminated between and among groups.
Results: Each of 6 DFAs identified 2–6 of the 48 predictor variables. These variables achieved 84%–100% classification accuracy for group membership.
Conclusions: Dysarthrias can be characterized by quantifiable temporal patterns in acoustic output. Because EMS analysis is automated and requires no editing or linguistic assumptions, it shows promise as a clinical and research tool.

KEY WORDS: dysarthria, classification, envelope modulation spectra

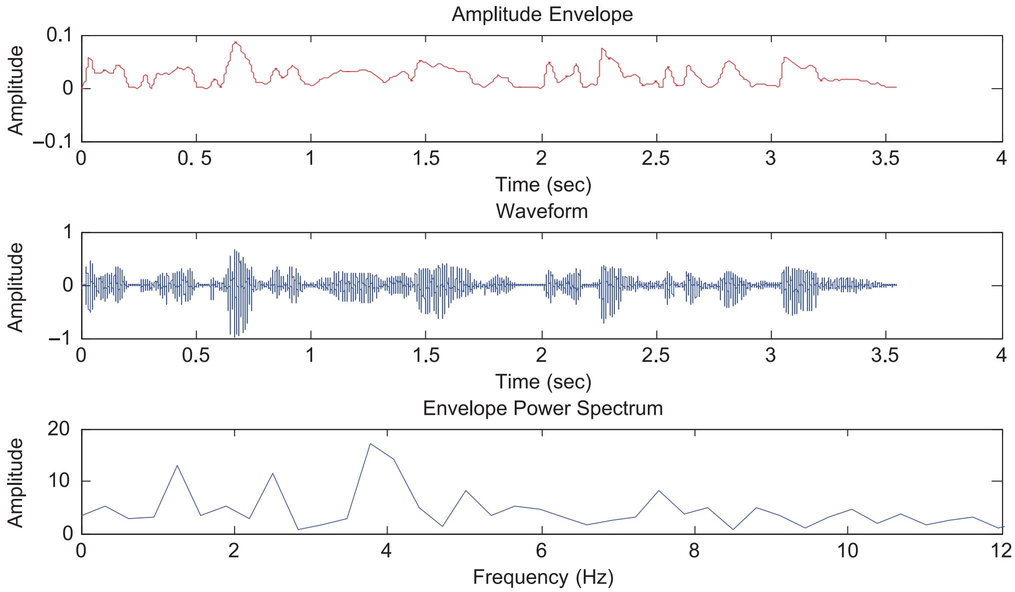
The dysarthrias are characterized, in part, by the ways in which the neurological defects interfere with outward flow of speech, resulting in the percept of disturbed speech rhythm. In a previous study, we applied segmental duration metrics developed for the classification of rhythms associated with different languages to dysarthric and healthy speech (Liss et al., 2009). These standard rhythm metrics are derived from an acoustically based segmentation of the speech signal into vocalic and consonantal intervals, and they are designed to capture differences in speech rhythm between and within languages with high versus low temporal stress contrasts—so called “stress timed” and “syllable timed” languages, respectively (Dellwo, 2006; Grabe & Low, 2002; Low, Grabe, & Nolan, 2000; Ramus, Nespor, & Mehler, 1999). Acoustic measures of vocalic and consonantal segment durations were obtained for speech samples from healthy individuals and those with hypokinetic, hyperkinetic, flaccid–spastic, and ataxic dysarthrias. Segment durations were used to calculate standard and new rhythm metrics. Stepwise discriminant function analyses (DFAs) were used to determine which sets of predictor variables (rhythm metrics) best discriminated between groups (control vs. dysarthrias; and among the four dysarthrias). DFAs pitting each dysarthria group against the combined others resulted in unique constellations of predictor variables that yielded equally high levels of classification accuracy, all exceeding 80%. These variables coincided in interpretable ways with perceptual features and underlying production constraints associated with the diagnostic categories (Kent & Kim, 2003).

These initial results were exciting as they suggested that segmental rhythm metrics can quantify rhythm disruptions peculiar to dysarthric subtypes, and map temporal disturbance patterns back to particular production deficits. However, there are a number of issues that constrain the wide application of these rhythm metrics to motor speech disorders. First, the manual marking of vocalic and consonantal intervals using a cursor function on a spectrographic display can be a challenge with dysarthric speech. Segmentation is based on operational definitions derived from healthy speech for visible acoustic landmarks associated with sound onset and offset (Peterson & Lehiste, 1960; White & Mattys, 2007). Dealing with degraded or absent landmarks requires experience and expertise in acoustic analysis to achieve reliable segmentation measures. Second, silent pauses within the connected speech require identification and an additional procedure (see White & Mattys, 2007). This further complicates the segmentation, particularly for dysarthrias that occasion frequent pauses between words and syllables due to impaired respiratory support. Clearly these problems are not insurmountable; however, they do render the possibility of an automated segmentation program less likely. Finally, these rhythm metrics carry with them fundamental linguistic assumptions about vocalic and intervocalic segments as meaningful entities in the perception of rhythm, which may or may not apply to degraded speech.

Here we report an alternative method for assessing speech rhythm that overcomes the problems of rhythm metrics but retains their benefits. This method involves the automated measurement of the temporal regularities in the amplitude envelope of the speech waveform, and requires no segmentation, no special procedures for silent pauses, and no linguistic assumptions. The envelope modulation spectra (EMS) is a spectral analysis of the low-rate amplitude modulations of the envelope for the entire speech signal and within select frequency bands. In Figure 1, the top graph shows the amplitude envelope of a speech waveform that is displayed in the middle graph. The envelope is obtained by half rectification of the signal and then by applying a low-pass filter with a cutoff of 30 Hz. The resulting envelope includes temporal variations in amplitude such as those that correspond to syllables. Regular durational patterns such as stressed–unstressed rhythms are also apparent. We can quantify these regularities by performing a Fourier analysis on the envelope resulting in a representation of the amplitude modulation rates that predominate in the signal. Below the waveform and envelope is a graph of the power spectrum of the envelope for this signal. Note that the peak energy (in decibels) is at 4 Hz, with other peaks present at higher and lower rates.

Tilsen and Johnson (2008) recently suggested that the modulation spectrum calculated for an envelope extracted

Figure 1. EMS dependent measures. The top graph shows the amplitude envelope of a speech waveform that is displayed in the middle graph (a male saying, "Auditory cognitive neuroscience is fun for the whole family"). The bottom graph is of the envelope modulation spectrum for this signal. (The waveform is normalized in amplitude between 0 and 1, and the spectrum of the down-sampled envelope of this waveform is presented in dB.)



from the frequency band of 700 Hz–1300 Hz (i.e., the signal is band-pass filtered and then an envelope extracted) may be useful to distinguish languages by speech rhythm. They also demonstrated that the number of segmental deletions in a particular sample was predictable from the energy in the spectrum at different rates (lower energy between 2 Hz and 5 Hz was associated with more deletions). This raises the prospect that the EMS may be a reliable predictor of speech production disruptions.

Tilsen and Johnson (2008) concentrated on the region between 700 Hz and 1300 Hz based on previous work by Cummins and Port (1998) on determining perceptual units of rhythmic stress in English (p centers). This frequency region would not be contaminated by f0 fluctuations or fricative noise and would be dominated by the onset and offset of vowels. Although these may be preferable conditions for determining language stress patterns, they are unnecessarily restrictive for examining rhythmic patterns in disordered speech. In the EMS, amplitude spectra are computed for the entire signal and for octave bands with center frequencies ranging from 125 Hz to 8000 Hz. This allows one to examine rhythmic patterns that are due to vowel nuclei, voicing, bursts and fricatives, and so forth. Crouzet and Ainsworth (2001) demonstrated that amplitude envelopes extracted from different frequency bands for normal speech signals are only partially correlated (see also Houtgast & Steeneken, 1985; Plomp, 1983). Thus, there is potentially independent information to be extracted from the envelopes of different frequency bands.

Although the EMS provides a detailed description of the low-rate amplitude modulations of the signal, one needs to derive a finite set of predictor variables in order to attempt to classify the dysarthric subtypes. In the present study, six variables were computed for each octave band as well as the entire signal (see Table 1). Two of these variables, peak frequency and peak amplitude, are measures of the dominant modulation rates. Peak frequency is the frequency of the peak with the largest amplitude, and peak amplitude is the amplitude of this peak normalized to the total energy in the spectrum (up to 10 Hz). These measures indicate the dominant rate and how dominant it is, respectively. The third variable is the amount of normalized energy in the region of 3 Hz–6 Hz (E3–6). These rates correspond to periods from 167 to 333 ms, which captures the majority of syllable durations in normal productions of English and Japanese (Arai & Greenberg, 1997). This region also includes the 4-Hz rate (250-ms period), which has been considered the dominant rate for normal speech (computed across the entire signal; Greenberg, Arai, & Grant, 2006; Houtgast & Steeneken, 1985) and has been shown to be related to the intelligibility of speech (Drullman, Festen, & Plomp, 1994; Houtgast & Steeneken, 1985). The last three predictor variables are also related to the 4-Hz rate. In pilot analyses with dysarthric and normal speech samples, the amount of energy in the energy power source (normalized for overall energy in the power spectrum) was computed for 0.5 Hz–Hz bands from 0 Hz to 10 Hz. Correlation computations revealed that energy in the bands below 4 Hz tended to be intercorrelated as were the bands above 4 Hz, but there was much less cross-correlation for pairs of bands that crossed the 4 Hz boundary. As a result, the spectral energy below 4 Hz (Below4) and the spectral energy above 4 Hz (Above4) were computed as separate

Table 1. The six dependent variables obtained for each of the octave bands and full signal in the EMS.

Measure	Description
Peak frequency	The frequency of the peak in the spectrum with the greatest amplitude. The period of this frequency is the duration of the predominant repeating amplitude pattern.
Peak amplitude	The amplitude of the peak described above (divided by overall amplitude of the energy in the spectrum). This is a measure of how much the rhythm is dominated by a single frequency.
E3–6	Energy in the region of 3–6 Hz (divided by overall amplitude of spectrum). This is roughly the region of the spectrum, around 4 Hz, that has been correlated with intelligibility (Houtgast & Steeneken, 1985) and inversely correlated with segmental deletions (Tilson & Johnson, 2008).
Below4	Energy in spectrum from 0–4 Hz (divided by overall amplitude of spectrum). The spectrum was split at 4 Hz, because it has come up as an important rate in previous work and because pilot work demonstrated that the amount of energy below and above 4 Hz was relatively uncorrelated across a variety of speakers and sentences.
Above4	Energy in spectrum from 4–10 Hz (divided by overall amplitude of spectrum).
Ratio4	Below4/Above4.

Note. EMS = envelope modulation spectra.

variables. These variables were computed by summing the normalized energy of 0.5-Hz bands from 0 Hz–4 Hz and 4 Hz–10 Hz, respectively. We chose 10 Hz (100-ms period) as the upper cutoff for the Above4 variable (as opposed to using the entire power spectrum) because we wanted to focus on suprasegmental variations in rhythm. The ratio of energy below 4 and above 4 Hz (Ratio4) was also included as a variable.

Method

Speakers

Forty-three speakers selected from a pool for a larger study provided speech samples that were analyzed for the current investigation: 12 with a diagnosis of ataxic dysarthria secondary to various neurodegenerative diseases (ataxic); 8 with hypokinetic dysarthria secondary to idiopathic Parkinson's disease (PD); 4 with hyperkinetic dysarthria secondary to Huntington's disease (HD); 10 with a mixed spastic–flaccid dysarthria secondary to amyotrophic lateral sclerosis (ALS); and 9 neurologically normal speakers (control). The speakers with dysarthria were selected because their speech deficits were of at least moderate severity (as per intelligibility measures conducted for the larger investigation), and because their perceived symptoms coincided to varying degrees with the cardinal speech features associated with the corresponding speech diagnosis (see Table 2). The presence of the cardinal speech features was established as part of the research protocol, which involved independent perceptual assessment by at least two certified speech-language pathologists.

Speech Stimuli

Productions of five sentences (see Appendix) were recorded as part of the larger investigation. Participants were fitted with a head-mounted microphone (Plantronics DSP-100), seated in a sound-attenuating booth, and read stimuli from visual prompts on a computer monitor. Recordings were made using a custom script in TF32 (Milenkovic, 2004; 16-bit, 44 kHz) and were saved directly to disc for subsequent editing and manipulation using commercially available software (SoundForge, 2004). The sentences were adapted by White and Mattys (2007) from a larger corpus (Nazzi, Bertoncini, & Mehler, 1998) for use in the study of speech rhythm differences in languages. These speech tokens are the same ones used in Liss et al. (2009), Set 1, allowing for a direct comparison of results between the present EMS analysis and the previous rhythm metrics analysis.

EMS

The present study calculated the modulation spectra for amplitude envelopes extracted from the full signal and seven octave bands (center frequencies of 125, 250, 500, 1000, 2000, 4000, and 8000). From each of these eight modulation spectra, six variables were computed (see Table 1). These 48 dependent variables (8 envelopes × 6 metrics) can be calculated from any signal using a fully automated program developed in MATLAB (Mathworks). The signal is filtered into the octave bands (pass-band eighth-order Chebyshev digital filters), and the amplitude envelope is extracted (half-wave rectified, followed by 30-Hz low-pass fourth-order Butterworth filter) and downsampled (to 80 Hz, mean subtracted). The power spectrum of each down-sampled envelope is calculated with a 512-point fast Fourier transform using a Tukey window and converted to decibels for frequencies up to 10 Hz (normalized to maximum autocorrelation). The six EMS metrics are then computed from the resulting spectrum for each band (and the full signal).

For each speaker, the average values obtained for the five sentences were calculated for all 48 variables.

Table 2. Speech features by dysarthria type.

Speaker group	Cardinal perceptual symptoms present to varying degrees in all speakers with dysarthria
ALS (F = 6, M = 4)	Prolonged syllables; slow articulation rate, imprecise articulation; hypernasality; strained strangled vocal quality
Ataxic (F = 6, M = 6)	"Scanning" speech; imprecise articulation with irregular articulatory breakdown; irregular pitch and loudness changes
PD (F = 2, M = 6)	Rapid articulation rate; rushes of speech; imprecise articulation; monopitch; reduced loudness; breathy voice
HD (F = 0, M = 4)	Irregular pitch and loudness changes; irregular rate changes across syllable strings
Control (F = 4, M = 5)	

Note. ALS = amyotrophic lateral sclerosis; PD = Parkinson's disease; HD = Huntington's disease; F = female; M = male. Blank cell indicates "not applicable."

Our initial analyses demonstrated that the EMS variables are only moderately correlated across neighboring frequency bands, and the correlations decreased (and became nonsignificant in some cases) as distance increased between frequency bands. These results match previous demonstrations by Crouzet and Ainsworth (2001) on the decreasing correlation between envelopes across increasingly distant frequency bands for speech. The variables obtained for the individual bands were also not redundant with those obtained for the full signal.

Discriminant Function Analysis

The goal of data analysis was to identify whether dependent variables derived from EMS could robustly distinguish speakers with dysarthria from healthy control speakers, and, furthermore, the extent to which these metrics could distinguish among the four different forms of dysarthria. A series of six stepwise DFAs was undertaken (SPSS, Version 15.0) to determine which sets of the 48 predictor variables best discriminate among speaker groups. At each stage of the DFA, the variable that minimized Wilks's lambda was entered, provided the F statistic for the change was significant ($p < .05$). At any point during the analysis, variables were removed from the DFA if they were found to no longer significantly lower Wilks's lambda ($p > .10$) when a new variable was added. Canonical functions, representing linear combinations of the selected predictor variables, were constructed by the DFA and used to create classification rules for group membership. The accuracy with which these rules classify the members of the group is expressed as a percentage correct. As a test of the robustness of the classification rules, we used cross-validation (also called the "leave-one-out") method. By this method, the DFA constructs the classification rules using all but one of the speakers, and then the excluded speaker is classified based on the functions derived from all other speakers.

Results

Six DFAs were conducted to evaluate the ability of predictor variables to discriminate (Analysis 1) control versus dysarthric; (Analysis 2) among all five speaker groups; and (Analyses 3–6) each dysarthria type against the combined others. In each case, the entire set of 48 dependent variables was entered as input to the stepwise DFA; the analysis identified the best set of predictor variables for maximizing the distances among group distributions; and these variables were entered into the classification function. Predictor variables in their order of importance, classification accuracy, cross-validation results, and misclassifications are reported for each analysis. In addition, we compare the present results to those obtained from the previous segmental duration metrics analysis (Liss et al., 2009). The predictor variables are named by the measure name, followed by the frequency band from which the measure was obtained; for example, Above4_1000 (energy above 4 Hz from the envelope spectrum obtained from the 1000-Hz central-frequency octave band), E3–6_Full, and Peak Amplitude_250. For the five binary classifications (Analyses 1 and 3–5), the mean level of each variable (higher or lower values) associated with the smaller classification set (e.g., controls in Analysis 1) is presented after each variable.

Analysis 1: Control Versus Dysarthria

This analysis classified speakers as either members of the control (9 members) or the dysarthria (34 members) group. Six of the 48 EMS metrics emerged as predictor variables: Above4_1000 (higher mean for control group); E3–6_125 (lower); E3–6_2000 (higher); Above4_4000 (higher); Ratio4_1000 (lower); and Above4_250 (higher). The classification function derived from the DFA accurately classified 100% of the speakers as members of control or dysarthric (with 95.3% accuracy on cross-validation). That is, these EMS variables were robustly capable of classifying sets of speech signals as coming from a healthy speaker or a speaker with dysarthria.

Analysis 2: All Five Speaker Groups

This analysis classified individuals as belonging to one of the five speaker groups (control, ataxic, ALS, HD, or PD). A different constellation of six variables emerged as predictive: Below4_8000; Above4_1000; E3-6_1000; Peak Amplitude_250; Peak Frequency_250; and Ratio4_250. The classification function derived from the DFA accurately classified 84% of the speakers as members of their designated group (with 67.4% on cross-validation). This compares favorably with the 79% classification accuracy (with four variables) obtained in the Liss et al. (2009) segmental duration metric study.

There were no misclassifications of control or HD speakers. The six misclassifications that did occur are as follows: 1 PD speaker as control, 1 ALS speaker as ataxic, and 4 ataxic speakers as ALS. In the previous study, 12 speakers were misclassified, and only 1 speaker (ALS) overlapped with the current DFA (see Table 3).

Analysis 3: PD Versus All Other Dysarthrias

In the remaining four analyses, only data from dysarthric speakers were included. The intent of these analyses was to see whether a particular subtype could be correctly categorized compared with members of the other three subtypes used in this study. Numbers of correct and incorrect classifications for each subtype are presented in Table 4. These analyses match those performed in Liss et al. (2009). In Analysis 3, the sentence sets were

Table 3. Classification and misclassification of group membership for Analysis 2.

	Predicted group membership					
Analysis/Speaker group	**Ataxic**	**ALS**	**HD**	**PD**	**Control**	**Total**
Ataxic	8	4				12
ALS	1	9				10
HD			4			4
PD				7	1	8
Control					9	9

designated as produced by speakers with hypokinetic dysarthria (8 members) or by speakers with one of the other dysarthria subtypes (26 members). Only three EMS variables were required to accurately classify all of the dysarthric speakers according to this classification scheme: Below4_8000 (lower); E3–6Hz_2000 (higher); and Ratio4_500 (lower). Cross-validation results were also at 100% accuracy. This same level of accuracy was achieved by the segment duration metrics study with three variables and no misclassifications.

Analysis 4: Ataxic Versus All Other Dysarthrias

In pitting ataxic versus the combined other dysarthrias, three variables were identified by the analysis: Below4_Full (higher); Peak Frequency_500 (higher); and Below4_1000 (higher). These variables allowed for 85.3% classification accuracy (with 79.4% accuracy on cross-validation). One ataxic speaker was misclassified as other dysarthria; and 3 ALS speakers and 1 PD were misclassified as ataxic. This level of accuracy compares very favorably with that of the segmental duration metrics study, in which 6 speakers were misclassified using three variables. One of these speakers (ALS) overlaps with the current DFA misclassification results.

Table 4. Correct and misclassification for Analyses 3–6.

Analysis	Speaker group	PD	Ataxic	ALS	HD
Analysis 3	PD	**8**	0	0	0
		0	**12**	**10**	**4**
Analysis 4	Ataxic	1	**11**	3	0
	Other dysarthria	**7**	1	**7**	**4**
Analysis 5	ALS	0	3	**9**	0
	Other dysarthria	**8**	**9**	3	**4**
Analysis 6	HD	1	1	2	**4**
	Other dysarthria	**7**	**11**	**10**	**0**

Note. Boldfaced numbers represent correct responses.

Analysis 5: Flaccid–Spastic (ALS) Versus All Other Dysarthrias

In this analysis, speakers were classified as belonging to either the ALS or the other dysarthria group. Two variables emerged—Peak Amplitude_8000 (higher) and Above4_1000 (lower)—yielding 82.4% classification accuracy (with 79.4% accuracy on cross-validation). Three ALS speakers were misclassified as other, and 3 ataxic speakers were misclassified as ALS. This level of accuracy is identical to the segmental duration metrics study (from one variable), with 2 misclassified speakers in common (ataxic).

Analysis 6: Hyperkinetic (HD) Versus All Other Dysarthrias

Finally, comparing HD with the other dysarthrias, two variables were identified: Peak Amplitude_250 (higher) and Above4_250 (lower). These variables achieved 88.2% classification accuracy with no misclassification of HD speakers (with 82.4% on cross-validation). Misclassifications as HD included 1 ataxic, 2 ALS, and 1 PD speaker. In comparison, the segmental duration metrics classification accuracy was 85% from one variable, and there was no overlap of misclassified individuals across the two analyses.

Discussion

The present study examined a methodological alternative to using segmental duration metrics to assess rhythm disturbances in the dysarthrias. Instead of manual segmentation of speech into vocalic and intervocalic intervals for duration calculations, a set of predictor variables was computed from the amplitude spectra of the full speech signal and seven octave bands. In all cases, the classification accuracy obtained by these EMS predictor variables matched or exceeded that of the segmental duration metrics used by Liss et al. (2009) for these particular speech tokens and comparisons. These preliminary results provide support for the development of EMS analysis in the study of speech production disorders. Because the analysis is automated, it also holds promise for wide clinical applicability to the extent it can sensitively capture changes in speech related to treatment or disease progression. In addition to the capability to be automated, the EMS measures have the benefit of being able to handle nonlinguistic parts of the signal, such as silences or noises, without special considerations. This is particularly important for the deployment of a rhythm measure for pathological speech.

One benefit enjoyed by the standard segmental duration metrics and not the EMS-dependent variables is relative interpretability in terms of traditional phonetic–suprasegmental constructs. This is because the segmental metrics all derive from durational relations among vocalic and intervocalic intervals with the express purpose of capturing temporal variability. The EMS is a substantially more complex representation of speech rhythm and is therefore potentially more powerful. However, at this point, it is not evident which rhythmic components of speech are represented within the various frequency bands of the EMS, or even whether stable correlates can be ascertained. An important goal going forward is to determine whether functional/perceptual significance can be assigned to amplitude modulations within specific frequency bands. A systematic study of the relationship between speech production, signal acoustics, and EMS measures has to be undertaken before we can offer functional analyses of differences in the EMS variables.

One possible adjustment to the computation of the EMS that may arise from systematic follow-up is changing the frequency bands from which the envelopes are extracted. In the present investigation, octave bands were utilized. One could base the bandwidths on psychoacoustically or physiologically derived scales such as equivalent rectangular bands (Moore & Glasberg, 1996) or equal distances on the basilar membrane (Greenwood, 1961). These adjustments may better align the EMS measures with perceptual representations of speech rhythm. Whatever the choice of bandwidth, it is important to sample the amplitude envelopes from across the frequency space. Across the six DFA analyses, significant predictor variables came from every one of the frequency bands. One surprising result was how much discriminative information was in the highest frequency band (8000-Hz center frequency and 5680–11,360 Hz bandwidth). Variables from this frequency region repeatedly show up as important in the DFA analyses. For example, the Below4 in the 8000-Hz band is the most discriminative variable for the classification of the PD group from the other dysarthrias. Interestingly, previous work on classifying voice disorders versus normal production has found that energy above 5 kHz is informative (Shoji, Regenbogen, Daw Yu, & Blaugrund, 1991; Valencia, Mendoza, Mateo, & Carballo, 1994). The fact that variables in this frequency region show up in the DFA could be due to their inherent importance or because they are correlated with a number of other important variables. Nevertheless, one wonders whether metrics of normal speech rhythm that do not include this frequency region, such as Tilsen and Johnson (2008), are missing important information about the temporal regularities in the speech signal.

It is also possible that the EMS is encoding important information about rhythm that is not available in other measures such as the segmental duration metrics. The fact that there was little overlap in the misclassifications obtained here and in Liss et al. (2009) suggests that the measures are not completely redundant, despite similar classification accuracy. The relationship of these two variable sets was examined by calculating the pairwise correlations of all EMS and segment duration measures for the entire 43-speaker dataset. Nearly all EMS variables were significantly correlated ($p < .05$) with at least one of the segmental duration metrics used by Liss et al. (2009) and were usually correlated with several. Likewise, all of the segmental duration variables were significantly correlated with one or more EMS metrics. Table 5 lists the highest correlated EMS variable for each of the 11 metrics used by Liss et al. It is clear from this analysis that much of the information in the segmental duration metrics is captured in the EMS variables. The one exception is the variable VarcoC: the normalized standard deviation of consonantal intervals. Only one EMS variable—Peak Amplitude_500—was significantly correlated, and this correlation was small ($r = .320$). VarcoC was designed to measure how variable consonant intervals are in a language normalized against the variability arising from differences in speaking rate (Dellwo, 2006; White & Mattys, 2007). VarcoC was the most predictive variable for ataxia and HD in the DFA from Liss et al. (2009). The EMS DFA resulted in better classification, but there was very low overlap in the misclassifications (only one misclassification overlap), probably because EMS does not robustly encode the information in VarcoC. It is possible that consonant interval information is not as well represented in EMS because the periods of the consonant intervals tend to be shorter than for vowels and thus fall out of the range of rates on which the EMS variables were calculated. In the present study, EMS variables only included energy out to a rate of 10 Hz (100-ms period). Variables could be developed that extend beyond this range to incorporate more phonetic-level modulations.

One might note from Table 5 that the EMS measures that are ratios of the energy below and above 4 Hz appear to most closely resemble the set of segmental duration metrics. Also, two variables from the 8000 Hz band are highly correlated with segmental measures. In particular, Below4_8000 had a very high inverse correlation with the measure of articulation rate ($r = -.862$). That is, the number of syllables spoken per second is nearly perfectly correlated with low amounts of energy below 4 Hz in the envelope spectrum from a band that extends from 5680 to 11360 Hz. Given this relationship, it is not surprising that this EMS variable was the best predictor of PD speakers versus other dysarthric speakers, given the salient rushed articulation rate characteristic of this subtype. This begs the question: Is EMS nothing more than a complicated way to extract speaking rate? To address this, we included the articulation rate values from Liss et al. (2009) in each of our DFA analyses. When comparing control versus dysarthric

Table 5. EMS variable with the greatest correlation for each of the segmental duration metrics of rhyhm used in Liss et al. (2009).

Segmental duration variable	Variable description	Highest correlated EMS variable	Correlation
ΔV	Standard deviation of vocalic intervals	Ratio4_500	.787
ΔC	Standard deviation of consonantal intervals	Ratio4_500	.682
%V	Percentage of duration composed of vocalic intervals	Ratio4_2000	.683
VarcoV	ΔV divided by mean vocalic duration	Ratio4_2000	−.740
VarcoC	ΔC divided by mean consonantal duration	Peak Amplitude_500	.320
VarcoVC	Standard deviation of VC intervals normalized for mean VC duration	Ratio_Full	−.405
nPVI–V	Normalized pairwise difference in successive vocalic intervals	Peak Amplitude_8000	−.769
rPVI–C	Pairwise difference in successive consonantal intervals	Ratio4_500	.663
nPVI–VC	Normalized pairwise difference in successive VC intervals	Ratio4_Full	−.479
rPVI–VC	Pairwise difference in successive VC intervals	Ratio4_Full	−.405
Articulation rate	Number of syllables produced per second	Below4_8000	−.862

Note. Short descriptions are provided for each of the segmental duration metrics. For more details about the calculation of these variables, see Liss et al. (2009).

speakers, articulation rate does not emerge as an independent predictor variable in the analysis. When entered alone, articulation rate achieves 81.4% classification accuracy for control versus dysarthric speakers. However, the EMS variable Above4_1000 on its own achieves 90.7% accuracy (88.4% on cross-validation); thus, the best EMS variable exceeds articulation rate as a predictor variable. Articulation rate does emerge as a variable in the five-group analysis, and this improves classification accuracy beyond the EMS variables alone (84% vs. 90.7%). Articulation rate on its own achieves 51.2% classification accuracy for the five-group analysis, so the additional benefit of EMS suggests it is not redundant information. For the four individual dysarthrias, articulation rate emerged as a predictor variable only for PD, as may be predicted. However, additional EMS variables significantly changed Wilks's lambda above and beyond articulation rate, suggesting that EMS provides more than mere articulation rate.

Conclusion

The EMS provides a promising new metric for rhythmic disturbances in the dysarthrias. The EMS variables used in the present study can be computed automatically, and they provide the basis for remarkably good classification of speakers into control versus dysarthric groups, as well into the particular subtypes of dysarthria. As was the case for the Liss et al. (2009) study, the speech samples selected for these analyses represent rather ideal candidates for discriminability, given that they were selected because of their perceptual distinctiveness. It will be necessary to submit a much larger and more varied corpus of speech samples to EMS analysis to explore its sensitivity and specificity. This will also be required to establish any stable and predictable relationships between specific frequency bands and production or perceptual phenomena. The real value of EMS, however, may be that it can measure temporal regularities that do not arise specifically from linguistic structure (as the segmental duration metrics were designed to do). The EMS variables may be sensitive to individual perturbations within diagnostic subtypes and may even be able to predict intelligibility deficits and challenges for the listener. Previous work by Liss and colleagues (Liss, Spitzer, Caviness, Adler, & Edwards, 1998, 2000; Spitzer, Liss, & Mattys, 2007) has demonstrated that perturbations in rhythmic structure can result in errors in word segmentation by the listener and that the types of segmentation errors obtained are predictable from the specific type of perturbation. It is possible that the EMS may provide a means to quantify these types of perturbation and provide some additional structure for the explanatory bridge between motor disorder subtype and communicative outcomes.

Acknowledgments

This work was supported by a research grant from the National Institute on Deafness and Other Communication Disorders (Grant 5 R01 DC 6859) to the first author and (Grant 5 R01 DC 4674) to the third author. The authors are especially appreciative for those who supported and participated in this research, including volunteers from Mayo Clinic Arizona; Phil

Hardt and members of the Huntington's Disease Society of America, Arizona Affiliate; the Arizona Ataxia Support Group; Pamela Mathy; and K. Sivakumar and his patients at the Neuromuscular Research Center, Scottsdale, Arizona.

References

Arai, T., & Greenberg, S. (1997). The temporal properties of spoken Japanese are similar to those of English. *Proceedings of Eurospeech, Rhodes, Greece, 2,* 1011–1114.

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 Cues for Sound Analysis, Aalborg, Denmark.

Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics, 26,* 145–171.

Dellwo, V. (2006). Rhythm and speech rate: A variation coefficient for delta C. In P. Karnowski & I. Szigeti (Eds.), *Language and language processing: Proceedings of the 38th Linguistic Colloquium, Piliscsaba 2003* (pp. 231–241). Frankfurt, Germany: Peter Lang.

Drullman, R., Festen, J. M., & Plomp, R. (1994). Effect of temporal envelope smearing on speech reception. *Journal of the Acoustical Society of America, 95,* 1053–1064.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In N. Warner & C. Gussenhoven (Eds.), *Papers in laboratory phonology 7* (pp. 515–546). Berlin, Germany: Mouton de Gruyter.

Greenberg, S., Arai, T., & Grant, K. (2006). The role of temporal dynamics in understanding spoken language. In P. Divenyi, K. Vicsi, & G. Meyer (Eds.), *Dynamics of speech production and perception* (Vol. 374, pp. 171–193). Amsterdam, The Netherlands: IOS Press.

Greenwood, D. D. (1961). Critical bandwidth and the frequency coordinates of the basilar membrane. *Journal of the Acoustical Society of America, 33*(10), 1344–1356.

Houtgast, T., & Steeneken, J. M. (1985). A review of the mtf concept in room acoustics and its use for estimating speech intelligibility in auditoria. *Journal of the Acoustical Society of America, 77*(3), 1069–1077.

Kent, R. D., & Kim, Y. J. (2003). Toward an acoustic typology of motor speech disorders. *Clinical Linguistics & Phonetics, 17*(6), 427–445.

Liss, J. M., Spitzer, S., Caviness, J. N., Adler, C., & Edwards, B. (1998). Syllabic strength and lexical boundary decisions in the perception of hypokinetic dysarthric speech. *Journal of the Acoustical Society of America, 104*(4), 2457–2566.

Liss, J. M., Spitzer, S. M., Caviness, J. N., Adler, C., & Edwards, B. (2000). Lexical boundary error analysis in hypokinetic and ataxic dysarthria. *Journal of the Acoustical Society of America, 107*(6), 3415–3424.

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

Low, E. L., Grabe, E., & Nolan, F. (2000). Quantitative characterisations of speech rhythm: "Syllable-timing" in Singapore English. *Language and Speech, 43,* 377–401.

Milenkovic, P. (2004). TF32 [Computer software]. Madison, WI: University of Wisconsin—Madison, Department of Electrical and Computer Engineering.

Moore, B. C. J., & Glasberg, B. R. (1996). A revision of Zwicker's loudness model. *Acta Acustica, 82,* 335–345.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Towards an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 756–766.

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

Plomp, R. (1983). The role of modulation in hearing. In R. Klinke (Ed.), *Hearing: Physiological bases and psychophysics* (pp. 270–275). Heidelberg, Germany: Springer-Verlag.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition, 73*(3), 265–292.

Shoji, K., Regenbogen, E., Daw Yu, J., & Blaugrund, S. M. (1991). High-frequency components of normal voice. *Journal of Voice, 5*(1), 29–35.

Sound Forge [Software]. (2004). Middleton, WI: Sony Creative Software.

Spitzer, S., Liss, J., & Mattys, S. (2007). Acoustic cues to lexical segmentation: A study of resynthesized speech. *Journal of the Acoustical Society of America, 122*(6), 3678–3687.

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

Valencia, N., Mendoza, L., Mateo, R., & Carballo, G. (1994). High-frequency components of normal and dysphonic voices. *Journal of Voice, 8*(2), 157–162.

White, L., & Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics, 35*(4), 501–522.

Received June 18, 2009

Accepted February 18, 2010

DOI: 10.1044/1092-4388(2010/09-0121)

Contact author: Julie M. Liss, Motor Speech Disorders Laboratory, Arizona State University Coor, 870102, Tempe, AZ 85287. E-mail: julie.liss@asu.edu.

Appendix. Recorded sentences.

1. The supermarket chain shut down because of poor management.
2. Much more money must be donated to make this department succeed.
3. In this famous coffee shop they serve the best doughnuts in town.
4. The chairman decided to pave over the shopping center garden.
5. The standards committee met this afternoon in an open meeting.