Characterizing rhythm in dysarthric speech using the temporal envelope

Eugenia San Segundo[1], Jonathan Delgado[2] and Lei He[3]

[1]Department of Spanish Language & General Linguistics, UNED (Spain), [2]Dept. of Developmental & Educational Psychology, La Laguna University (Spain), [3]Department of Computational Linguistics, University of Zurich (Switzerland)

PREVIOUS STUDIES ON RHYTHM IN DYSARTHRIA

First descriptions of dysarthria: reduced articulation rate + irregular duration contrasts between stressed and unstressed syllables + scanning rhythm (*staccato*) [1, as cited in 2].

- But this description refers to one particular type of dysarthria: ataxic dysarthria.

Recent investigations on a wide range of dysarthrias show disparate results:

- In [3], acoustic measures of vocalic and consonantal segment durations allow to distinguish control speech from dysarthria and to discriminate dysarthria subtypes.
- In [4], none of the rhythm metrics based on segmental durations could differentiate disordered from healthy speakers.

MacNeilage's frame/content theory

This theory [5] unifies the different approaches to the study of rhythm, as explained in [6]:

- metrics quantifying the durational variability of vocalic and consonantal intervals focus on the *content* perspective.
- modulation-based approaches target the *frame* perspective with different emphases: recurring frequencies in the temporal envelope [7] and coordination between envelopes at slower and faster rates [8].

As content approaches to rhythm in dysarthria have resulted in disparate conclusions, we analyzed rhythm from the speech temporal envelope.

METHODS

Subjects: 30 Spanish speakers (15 dysarthric; 15 non-dysarthric)

Speech material: 4 phonetically balanced sentences

Acoustic analyses:

- Bandpass filtered (700-1300 Hz) signal to detect the P-centers.
- Filtered signal was full-wave rectified and downsampled to the Nyquist frequency of 20 Hz, yielding the temporal envelope.
- Calculated 5 spectral measures from the temporal envelope of each sentence: CENTROID, SPREAD , ROLLOFF, FLATNESS and ENTROPY.

Scan QR code for more info about the measures ↓

RESULTS

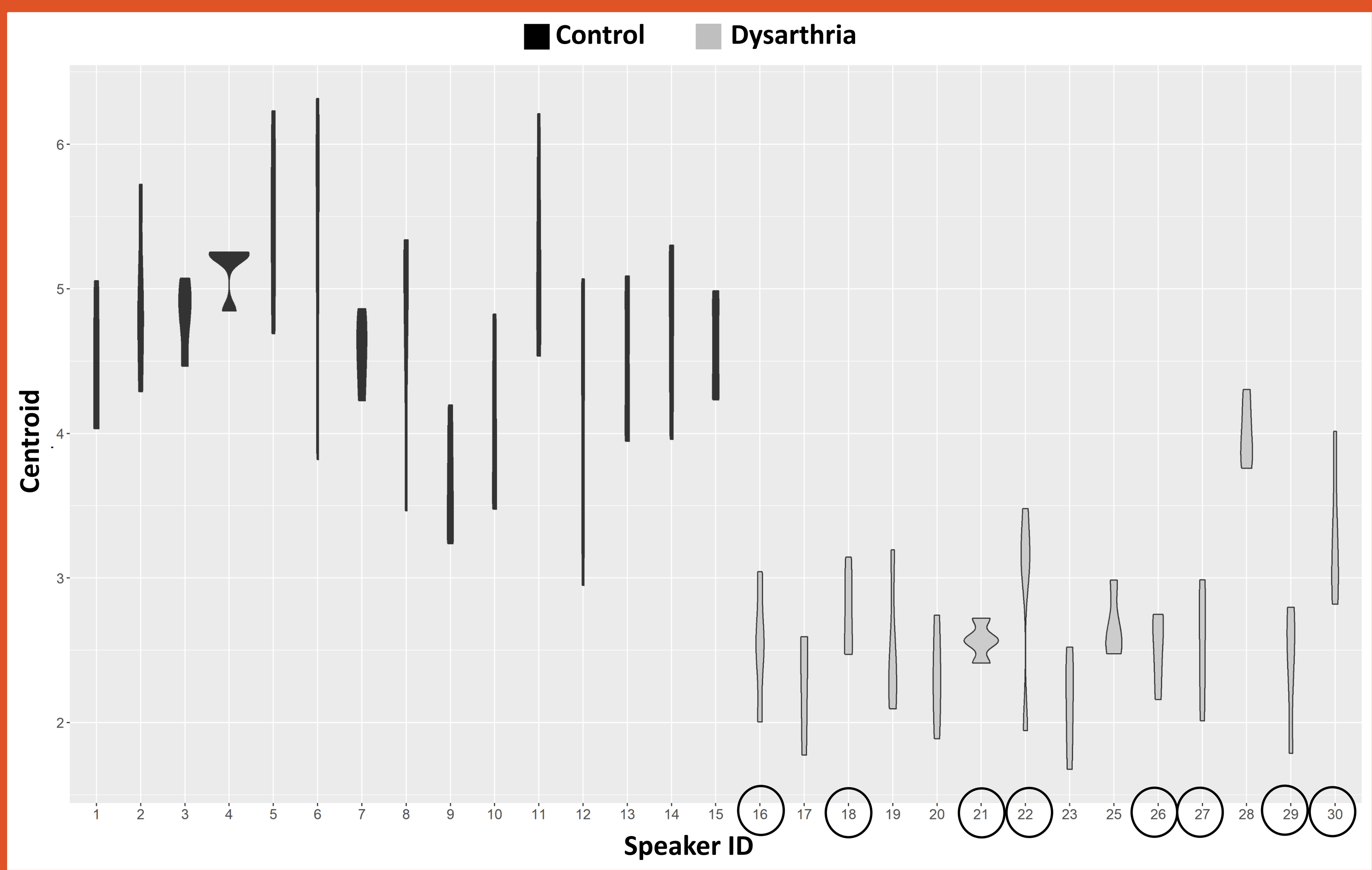
Predictor	Estimate	Std. Error	z
CENTROID	–4.51	1.72	–2.63 **
SPREAD	–5.95	2.84	–2.09 *
FLATNESS	10.87	11.34	0.96
ROLLOFF	0.22	0.73	0.30
ENTROPY	3.11	1.88	1.66 .

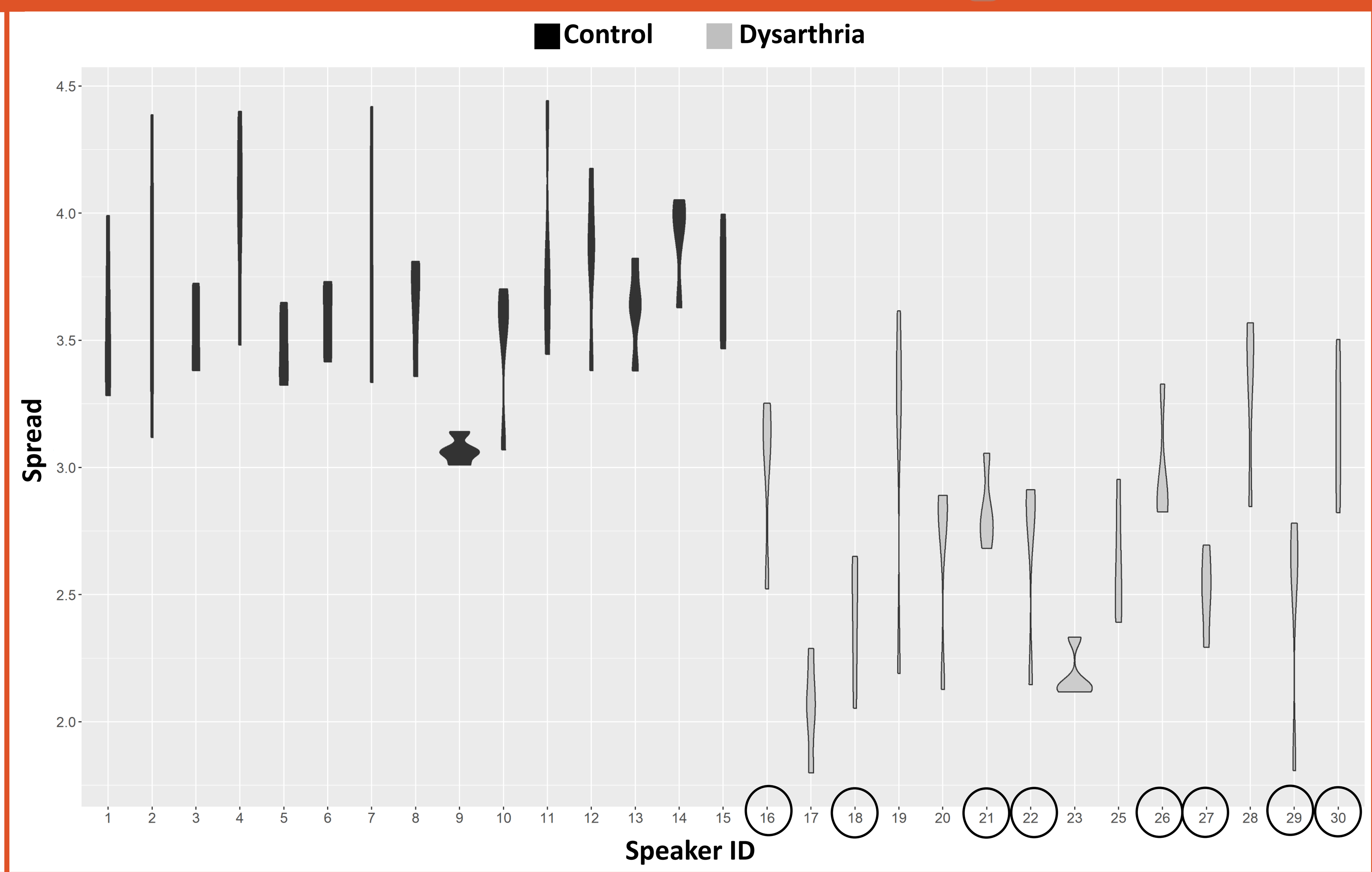
Table 1: Results of the fixed-effect predictors in the generalized linear mixed model (binomial logit). **p<0.01; *p<0.05; . p<1.

Dysarthric speakers present a significantly lower CENTROID…

…and significantly lower SPREAD in their speech temporal envelope

- The reference of the response variable group is C (control speakers).
- The model explains between 66.54% and 88.76% of the dependent variable variation (Cox & Snell's R^2 = 0.6654; Nagelkerke's R^2 = 0.8876).



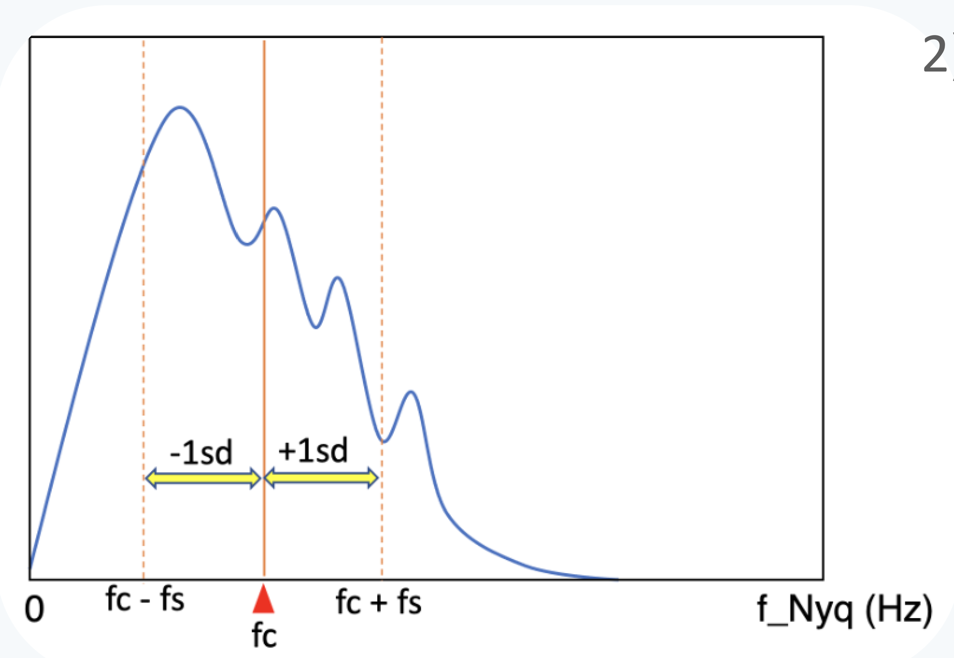


○ = speakers with ataxic dysarthria

DISCUSSION

1) Lower CENTROID freq. in dysarthric speakers:

- This agrees with our expectations:
  - Dysarthric speakers exhibit a slow speech rate due to slow articulatory movements
  - A stretched rhythmic frame was therefore expected, together with a centroid shifted towards lower frequencies



2) Narrower spectral SPREAD in dysarthric speakers:

- This suggests that the oscillations in the rhythmic frame were more regular around the centroid.
  - Possible explanation: jaw displacements in the non-dysarthric population are variable while jaw oscillations in dysarthria do not exhibit large variation. Instead, in terms of openness, both the jaw and the mouth remain stationary in these speakers throughout utterance production.
  - These results can be explained by the fact that dysarthria is a motor speech disorder in which the muscles used to produce speech are damaged, paralyzed, or weakened.

CONCLUSIONS

- Since previous dysarthric studies had shown that content-based approaches to rhythm (durational variability in different phonetic intervals) resulted in disparate findings, the method proposed here analyzes rhythm from the speech temporal envelope. It presents the **advantage of not requiring audio transcriptions or phonetic interval segmentation**.
- Previous L2 studies have shown that **rhythmic characteristics may not be sufficiently explained from the traditional content-based perspective alone** [6], highlighting the importance of the frame aspect in rhythm acquisition. In a similar way, investigations on speech disorders can benefit from this combined perspective.
- Since the sonorant/obstruent alternations typically follow the opening-closing mouth movements, according to the frame-content theory [5], **future dysarthric studies could analyze the spectral coherence between the temporal envelope and mouth opening-closing articulatory data**, as well as correlations with **jaw openness ratings in the VPA perceptual protocol**.
- Future studies will consider analyzing **ataxic, spastic and mixed dysarthric patients separately,** since voice impairment in ataxic dysarthria can be more irregular than in other dysarthrias [2].

[1] Holmes, G. 1917. The symptoms of acute cerebellar injuries due to gunshot injuries. *Brain*, 40(4), 461–535.
[2] Ziegler, W. 2016. The phonetic cerebellum: cerebellar involvement in speech sound production. In: Marien, P., Manto, M. (eds), *The linguistic cerebellum*. Academic Press, 1–32.
[3] Liss, J. M., White, L., Mattys, S. L., Lansford, K., Lotto, A. J., Spitzer, S. M., Caviness, J. 2009. Quantifying speech rhythm abnormalities in the dysarthrias, *J. Speech Lang. Hear. Res*. 52(5), 1334–1352.
[4] Lowit, A. 2014. Quantification of rhythm problems in disordered speech: A re-evaluation, *Philos. Trans. R. Soc. B.*, 369, 20130404.
[5] MacNeilage, P.F. 1998. The frame/content theory of evolution of speech production. *Behav. Brain Sci*. 21, 499–511.
[6] He, L. 2022. Characterizing first and second language rhythm in English using spectral coherence between temporal envelope and mouth opening-closing movements. *JASA* 152(1), 567–579.
[7] Tilsen, S., Johnson, K. 2008. Low-frequency Fourier analysis of speech rhythm. *JASA* 124 (2), EL34–EL39.
[8] Leong, V., Stone, M, Turner, R, Goswami, U. 2014. A role for amplitude modulation phase relationships in speech rhythm perception. *JASA* 136, 366-381.