Clustering approaches to dysarthria using spectral measures from the temporal envelope

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

[1]Phonetics Laboratory, Spanish National Research Council (CSIC), Madrid, Spain
[2]Department of Developmental & Educational Psychology, La Laguna University, Spain
[3]Department of Computational Linguistics, University of Zurich, Switzerland

Speech Prosody 2024

SP2024 Leiden

PREVIOUS STUDIES ON DYSARTHRIA

Dysarthria: speech disorder stemming from neurological factors that causes difficulties in speech motor programming and execution.

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 research 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.

RHYTHMIC APPROACHES

Approaches to speech rhythm are diverse, focusing on different aspects of the (semi-)regularities and variabilities in speech production (see [6] for a general overview).

The **frame/content theory** [5] unifies the different approaches to the study of rhythm:

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

METHODS

Subjects: 15 dysarthric; 15 non-dysarthric speakers.

Speech material: 4 phonetically balanced sentences (Spanish Matrix Sentences Test) read aloud.

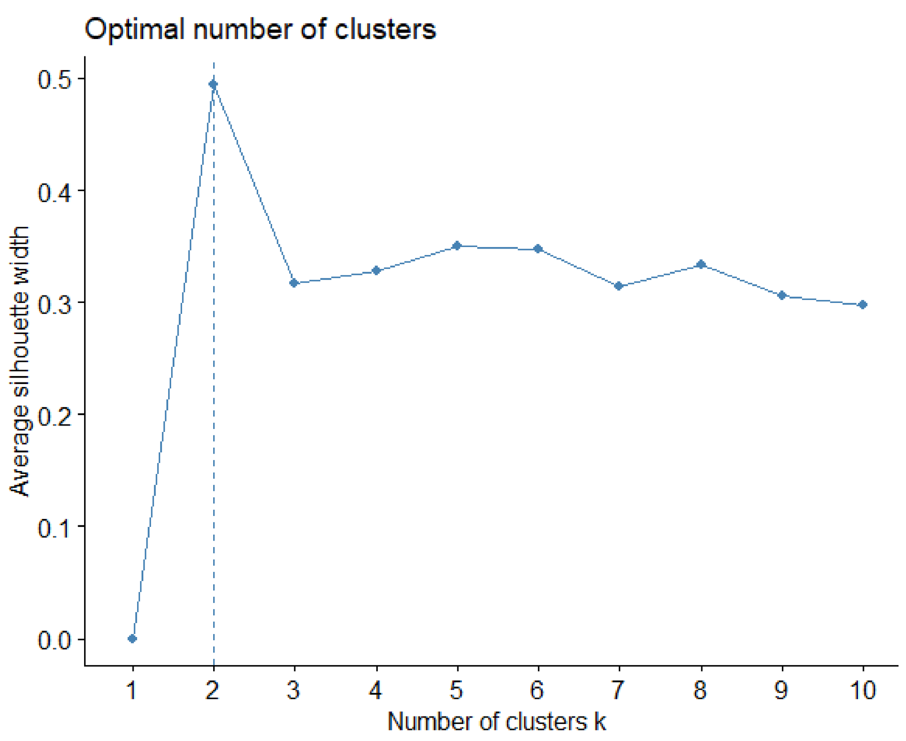
Acoustic analyses:

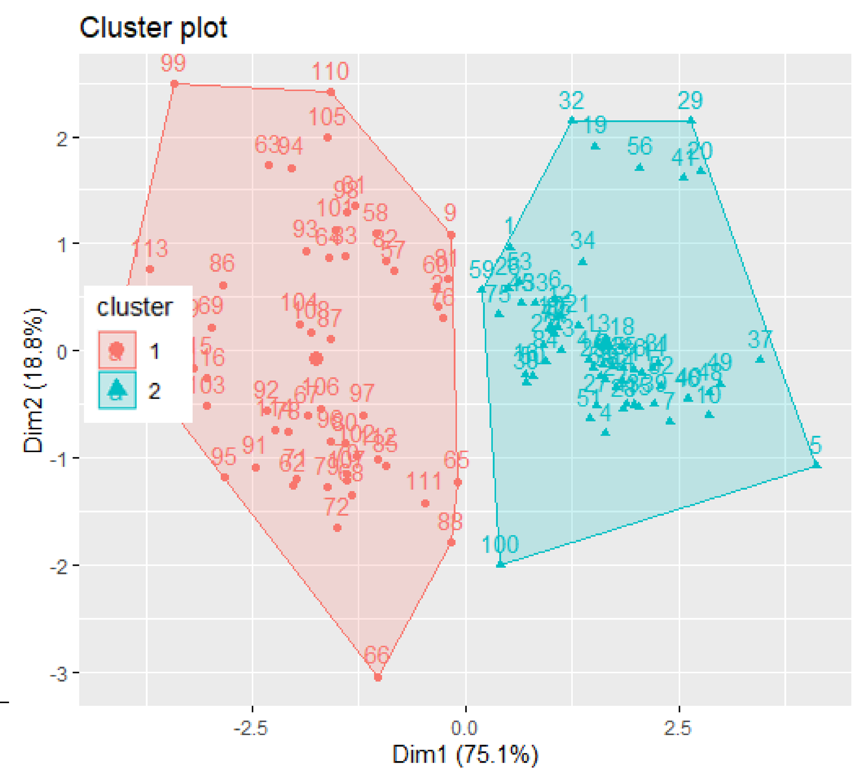
- Bandpass filtered (700-1300 Hz) signal to detect the P-centers.
- Filtered signal was full-wave rectified and downsampled to the Nyquist frequency of 20 Hz, yielding the temporal envelope.
- Calculated **5 spectral measures from the temporal envelope of each sentence:** CENTROID, SPREAD, ROLLOFF, FLATNESS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

Scan QR code for more info about the 5 spectral measures

RESULTS & DISCUSSION

1) K-means clustering





Cluster 1 matches well the 56 observations of dysarthric speakers (first 14 speakers * 4 sentences); i.e., datapoints 61 to 116.

Cluster 2 matches well the 60 observations of the control speakers (remaining 15 speakers * 4 sentences); i.e., datapoints 1 to 60.

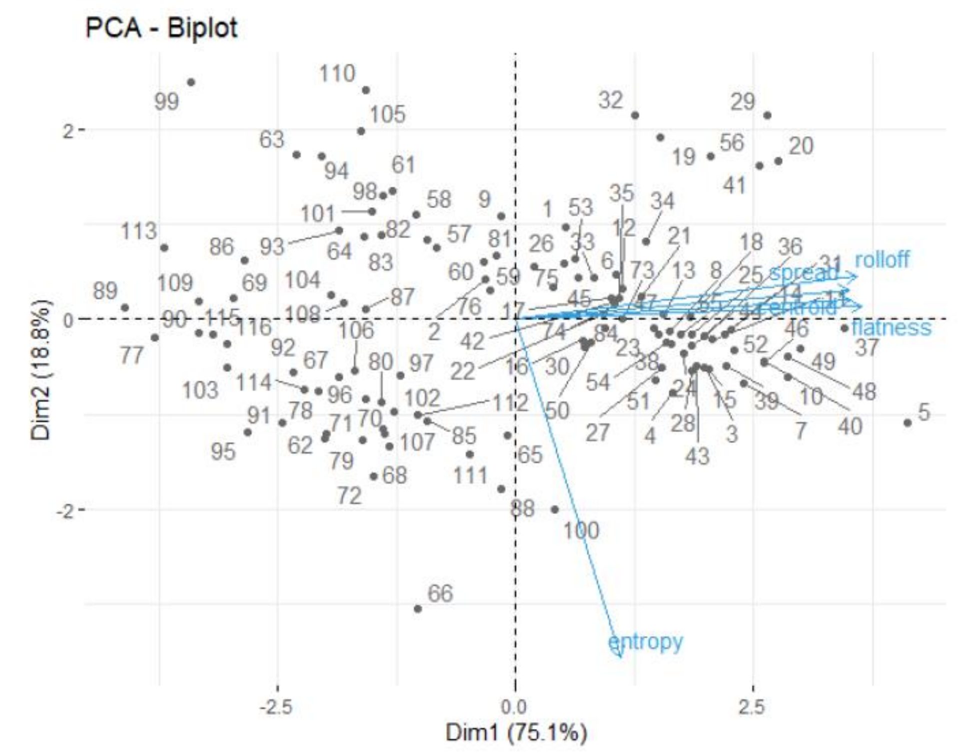
Cl.	Centroid	Spread	Flatness	Rollof	Entropy
1	-0.898	-0.838	-0.882	-0.888	-0.205
2	0.838	0.782	0.824	0.828	0.191

Cluster 1 is characterized by (-) values in the five spectral measures, while **Cluster 2** presents (+) values.

<u>**Dysarthric**</u> speakers present a stretched rhythmic frame and a CENTROID shifted towards lower frequencies, as well as a narrower spectral SPREAD in the temporal envelope, in comparison with control speakers.

2) Principal Component Analysis (PCA)

Dim.	Eigenvalue	Variance%	Cum. variance%
1	3.753	75.07	75.07
2	0.938	18.76	93.83
3	0.200	4.00	97.83
4	0.074	1.49	99.32
5	0.034	0.68	100

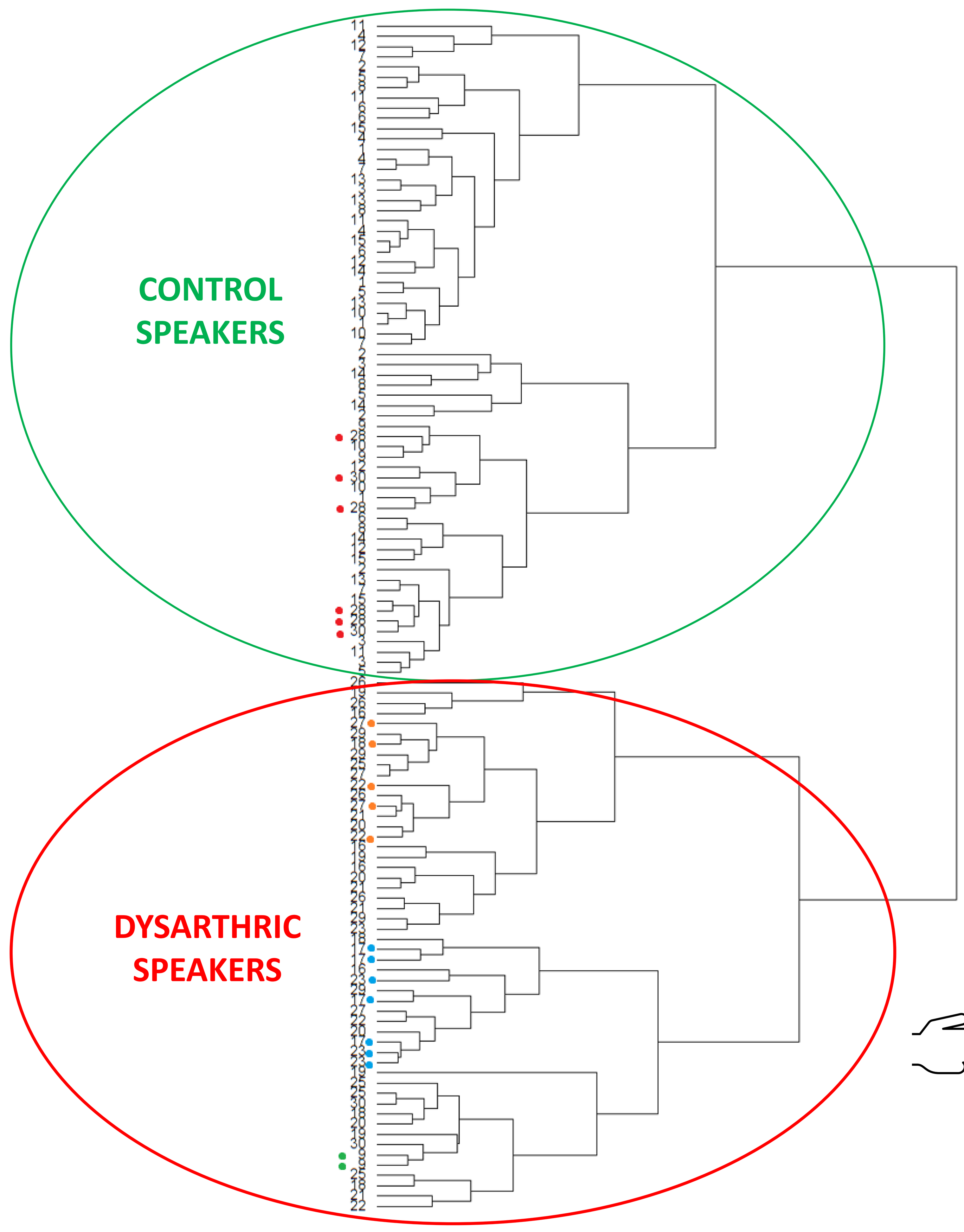


A PCA - biplot merges a PCA plot and a loadings plot:

- **PCA plot** shows clusters of samples based on their similarity
- **Loadings plot** shows how strongly each characteristic influences a principal component.

When vectors are close, forming a small angle, the variables that they represent are positively correlated (e.g. CENTROID, SPREAD, FLATNESS and ROLLOF). If they meet each other at 90°, they are not likely to be correlated (the 4 afore-mentioned variables and ENTROPY).

3) Hierarchical agglomerative clustering



Hierarchical clustering was conducted to visualize possible outliers and to find possible subgroups within each main cluster. The **dendrogram** show that:

- Only two sentences of Control Speaker #9 (see **green points**) were misclassified as dysarthric.
- All the sentences of Dysarthric Speaker #28 and two of Dysarthric Speaker #30 (see **red points**) were classified in the control group.

Besides, within the dysarthric group, at least two trends can be observed:

- Speakers #18, #22 and #27 cluster together (see **orange points**) → They all have ataxic dysarthria with a tumor diagnosis.
- Speakers #17 and #23 also cluster together (see **blue points**) → They both present flaccid-spastic dysarthria, with an ALS diagnosis.

Type of dysarthric speakers

Speaker	Sex	Age	Diagnosis	Dysarthria type
16	F	34	CVA	Ataxic
17	F	48	ALS	Spastic-flaccid
18	F	33	Tumor	Ataxic
19	M	21	CP	Spastic
20	F	30	CP	Spastic
21	F	51	CVA	Ataxic
22	F	40	Tumor	Ataxic
23	M	55	ALS	Spastic-flaccid
25	M	49	SCA-7	Spastic-ataxic
26	F	59	CBD	Ataxic
27	F	45	Tumor	Ataxic
28	M	39	CCT	Spastic-ataxic
29	F	47	CVA	Ataxic
30	F	52	Tumor	Ataxic

CVA: cerebrovascular accident; ALS: amyotrophic lateral sclerosis; CP: cerebral palsy; SCA-7: spino cerebellar ataxia-7; CBD: corticobasal degeneration; CCT: cranio-cerebral trauma

CONCLUSIONS

- We can conclude that the five spectral measures computed from the temporal envelope of read sentences seem to separate well between dysarthric and non-dysarthric speakers, using two different types of clustering techniques.
- However, more studies are needed to explore why two control speakers were classified as dysarthric: maybe idiosyncratic slow rate or muffled voice quality, etc.
- Likewise, it is necessary to explore which acoustic variables lie behind the clustering together of speakers with the same type of dysarthria:
    - ataxic dysarthria from a tumor diagnosis
    - spastic dysarthria from an ASL diagnosis

[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
[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.

Funding: **Grant PID2021-124995OA-I00 funded by MICIU/AEI/ 10.13039/501100011033 and by ERDF/EU.**

CSIC Consejo Superior de Investigaciones Científicas · Universidad de La Laguna · Universität Zürich UZH · Ministerio de Ciencia, Innovación y Universidades · Cofinanciado por la Unión Europea · Agencia Estatal de Investigación