On the use of low-level spectral features to distinguish speakers with Parkinson's disease from typical speakers.

Slis, A.[1], Fougeron, C.[1], Pernon, M.[1,2], & Lancia, L.[1]
[1]LPP, UMR 7018, CNRS/U. Sorbonne-Nouvelle, Paris, France.
[2]Department of Clinical Neurosciences, Geneva University Hospital, Switzerland.

Introduction. This study explores whether the acoustic output of speech productions provides enough information to distinguish speakers with Parkinson's disease (PD) from typical speakers. PD is a progressive neurogenetic disorder, caused by deficient cortico-basal ganglia-thalamic neural circuits, which are involved in the planning and execution of movements[1]. In over 80% of cases, people who are affected by PD develop Hypokinetic dysarthria[2], which is characterized, among other things, by slower and smaller articulatory movements with reduced peak velocities of articulators[3,5]. Although it is recognized that there is not a one-to-one relationship between the actual movements shaping the vocal tract and the resulting acoustic output, several acoustic features, such as formant transitions, have been linked back to changes in vocal tract shape. PD speakers, for example, show reduced F2 transitions which is said to reflect a reduced rate of change of vocal tract shape[6]. Although formant slope patterns thus shed light on some aspects of articulatory behavior, isolating formant curves is time consuming, susceptible to error[7], and possibly ignoring, yet unknown, important acoustic information caused by changes in vocal tract shape. This study explores the feasibility of considering the complete shape of the spectrum envelope and its patterns of change over time by extracting Mel-Frequency Cepstrum Coefficients (MFCC)[8,9], thus avoiding a priori assumptions about the important features in a spectrum. MFCCs have been shown to accurately represent the relevant perceptual characteristics of the speech signal[8] and their average squared change correlates consistently across speakers with dynamical changes in the vocal tract[9]. The hypothesis is that analysing the acoustic signal in terms of MFCCs provides sufficient spectral information to separate the speech productions of people diagnosed with Parkinson's Disease (PD) from typical speakers.

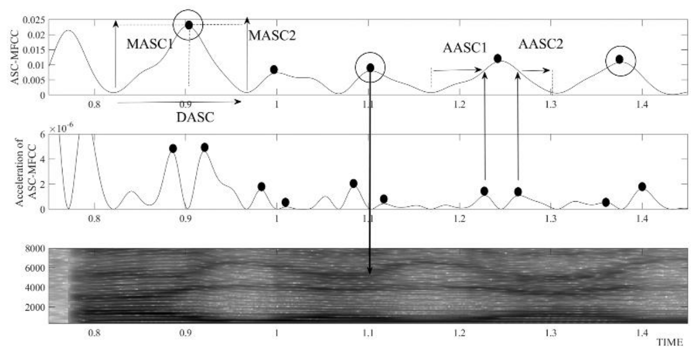

Figure 1: example of /ajajaj/ typical speaker; the curve of average squared changes of the MFCCs, the acceleration of average squared changes of the MFCCs and spectrogram. Only the circled peaks were selected for analysis: from vowel to glide in /aj/.

Methods. 13 French speaking PD speakers with dysarthria (severity: mild (6 speakers), moderate to severe (7 speakers)), and 13 French speaking typical speakers, matched in age and gender, produced the vowel-glide or glide-vowel sequences /ajajaj/, /ujujuj/ and /wiwiwi/ 4 times. Following the method described in[9], power spectra were first calculated from the acoustic waveform. Next, a MEL filter bank was applied to these power spectra and the summed spectral energy in each band was calculated. Finally, the DCT of the logarithmically transformed filter bank energies was calculated. To better focus on the consequences of changes in vocal tract configurations, DCT coefficients 2-13 were considered, discarding the F0. We then computed the squared differences between consecutive coefficient values and their average value (ASC_MFCC), resulting in a curve as exemplified in figure 1. This curve was used to extract the variables of interest with an analysis inspired by articulatory movement studies, in which the ASC_MFCC is annotated as an articulator velocity curve in terms of events. From each minimum ASC_MFCC (the steady state during the vowel) to the next maximum ASC_MFCC (the point of highest spectral energy change, i.e., during the transition to the glide), the peak value of ASC_MFCCs (MASC1) was computed. The same was done for each maximum to the next minimum (MASC2). DASC is the total duration of the vowel-glide gesture from minimum to minimum (i.e., from steady state vowel to steady state glide). In addition, the derivative was calculated from the ASC_MFCC curve (acceleration of the average squared change of the MFCCs). From this curve (see figure 1), the time from a minimum ASC_MFCC to a maximum acceleration peak (AASC1) and time from the maximum acceleration to a minimum ASC_MFCC valley were computed.

Results and discussion. MASC2, and AASC2 showed similar patterns as MASC1 and AASC1 and are not discussed any further. Because the method captures acoustic information with which the degree of change in vocal tract shape can be expressed, we expected that PD speakers show lower MASC values than typical speakers.

In addition, we expected that moderate to severe dysarthric patients show the largest differences in MASC.
It can be observed from figure 2 that for /ajajaj/, the measure MASC1 indeed separates PD speakers with both levels of severity from typical speakers effectively. /ujujuj/ and /wiwiwi/ showed lower MASC1 values for the moderate to severe group. Producing /ujujuj/ resulted in larger values for DASC in the moderate to severe group, but for /ajajaj/ and /wiwiwi/, the durational variable DASC did not show observable differences between typical and dysarthric speakers. Finally, the measures for AARC1 did not differ for the typical and the dysarthric speakers.

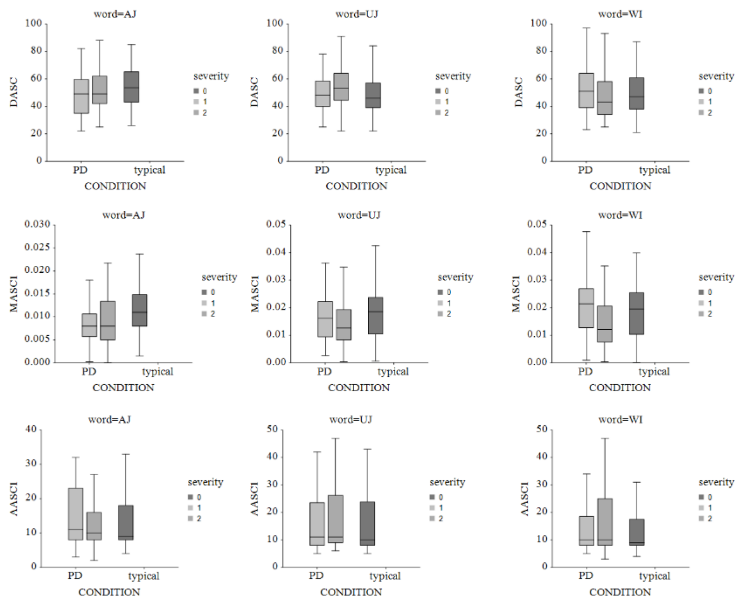

Figure 2 DASC (upper row), MASC1 (middle row) and AASC1 (lower row) for the three stimuli /ajajaj/ (left column), /ujujuj/ (middle column) and /wiwiwi/ (right column). Severity: 0: typical speakers; 1: mild; 2: moderate to severe.

The preliminary data show that using ASC_MFFCs as a tool to distinguishing typical from atypical PD speech is promising. The lower values for MASC1s characterizing the PD speakers suggest that the vocal tract changes are slower, which is consistent with the slower and smaller articulatory movements and reduced peak velocities of articulators, found in earlier kinematic studies[4,5,6], as well as the reduced F2 slopes[6]. The slightly different MASC1 results for /ajajaj/ compared to /wiwiwi/ and /ujujuj/ possibly originate from differences in articulatory behaviour. It is speculated that PD speakers likely face more difficulties in shaping the vocal tract to produce the contrasts of /ajajaj/ (the production of which is already affected in the mild dysarthric speakers), than /ujujuj/ and /wiwiwi/ (which show differences in the moderate and severe cases): the first stimulus requires a change from a low to a high position of the tongue and possible involvement of jaw movement, which is frequently affected in PD speakers[5,6], whereas the later two stimuli require a back-front or front-back movement of the tongue and more activation of the lips.

1. Duffy, J. R. (2005). *Motor speech disorders: Substrates, differential diagnosis, and management.* St. Louis, Mo: Elsevier Mosby.
2. Ramig, L.O., Fox, C., & Sapir, S. (2008). Speech treatment for Parkinson's disease. *Expert Rev. Neurotherapeutics 8*(2), 299–311
3. Forrest, K., & Weismer, G. (1995) Dynamic Aspects of Lower Lip Movement in Parkinsonian and Neurologically Normal Geriatric Speakers’ Production of Stress. *Journal of Speech and Hearing Research*, 0022-4685.

5) Yunusova, Y., Weismer, G., Westbury, J.R., Lindstrom, M.J. (2008). Articulatory movements during vowels in speakers with dysarthria and healthy controls. *Journal of Speech, Language, and Hearing Research, 51*, 596–611
6) Walsh, B. & Smith, A. (2012). Basic Parameters of Articulatory Movements and Acoustics in Individuals with Parkinson’s Disease. *Movement Disorders, 27*(7): 843–850. 357-366
7) Shadle, C.H., Nam, H., and Whalen, D.H. (2016). Comparing measurement errors for formants in synthetic and natural vowels. *The Journal of the Acoustical Society of America, 139*, 713; https://doi.org/10.1121/1.4940665
8) Davis, S. Mermelstein, P. (1980) *Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences*. In IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 28 No. 4, pp
9) Goldstein, L. (2019). The Role of Temporal Modulation in Sensorimotor Interaction. *Frontiers in Psychology, 06* | https://doi.org/10.3389/fpsyg.2019.02608.