Automatic analysis of the prosodic variations in Parkinsonian read and semi-spontaneous speech

Céline de Looze, Alain Ghio, Stefan Scherer, Gilles Pouchoulin, François Viallet

AUTOMATIC ANALYSIS OF THE PROSODIC VARIATIONS IN PARKINSONIAN READ AND SEMI-SPONTANEOUS SPEECH

Céline De Looze[1], Alain Ghio[2], Stefan Scherer[1], Gilles Pouchoulin[2], François Viallet[2]

[1]Speech Communication Lab, Trinity College Dublin, Ireland
[2]Laboratoire Parole et Langage, CNRS, Université d'Aix-Marseille,
Centre Hospitalier du Pays d'Aix, Aix-en-Provence, France
deloozec@tcd.ie; alain.ghio@lpl-aix.fr

Abstract

This work presents an automatic analysis of Parkinsonian speech prosody in read and spontaneous tasks. We first evaluate the reliability of existing algorithms for the automatic analysis of speech prosody and show that these tools are efficient for the analysis of Parkinsonian speech if a large amount of data is processed. We then investigate the prosodic characteristics of 25 control and 30 Parkinsonian subjects (in *off* (non-medicated) state) in both tasks and show that Parkinsonian read and spontaneous speech are characterized by longer pauses as well as smaller variations in pitch range, voice intensity level and articulation rate; these characteristics being all the more marked in semi-spontaneous speech. These results suggest that some linguistic functions, normally expressed via prosodic variations, may not be conveyed in Parkinsonian speech, particularly in a semi-spontaneous task.

Index Terms: prosodic variations, automatic measurements and tools, interpausal run, Parkinson disease, dysprosody

1. Introduction

Parkinsonian speech associated with hypokinetic dysarthria has often been described as monotonous, monoloudness, stress-reduced, variable speech rate and marked by inappropriate silences and paroxysmal acceleration [1-9]. While these descriptions were first based on perceptual assessments [1] [2], they were later augmented by instrumental analyses [7-9]. Acoustic analyses, which provide more objective measurements, have been generally considered as *"augmenting the impressionistic nature of perceptual description*" [10].

This approach is, however, not an easy task. It requires transcriptions and annotations at different levels of analysis (e.g. phonemes, syllables, prosodic units), which can be both time-consuming and error-prone when manually carried out and when processing large amounts of data. An alternative practical approach is to use automatic tools. This requires however that they are reliable for the analysis of pathological speech, with remaining inaccuracies being compensated by the large amount of data processed.

In terms of prosodic analyses, the definition of a temporal span or prosodic domain for the extraction of prosodic information is all the more crucial. While global measurements taken from entire recordings may provide a preliminary description on the prosodic characteristics of pathological speech, they cannot however capture the finer details of prosodic variations in speech. Yet, prosodic variations assume many communicative functions, e.g. give information about the hierarchical dimension and relational organization of discourse units and indicate a speaker's social states and attitudes [11].

In this study, we propose to use interpausal run (IR) as a unit for prosodic analysis as it can provide information on prosodic variations in speech and allows for an automatic analysis of large amounts of data. Acoustic analysis based on these interpausal run annotations specifically provides measurements of variations in pitch range, voice intensity and speech rate in order to further refine the current descriptions of Parkinsonian speech prosody. In this paper we first present an evaluation of three algorithms that we used for the automatic extraction of pitch extrema [11], syllable nuclei [12] and speech/pause intervals [13] in order to investigate their reliability for the study of pathological speech. We then report our automatic analysis on the prosodic variation characteristics of Parkinsonian speech, which compares the read and semi-spontaneous speech of 25 control and 30 Parkinson subjects (*off* state).

2. Corpora

2.1 *Selected subjects*

Thirty patients with Parkinson's disease (PD) (19 male and 11 female) were selected for this study. Speech recordings were carried out by expert phoneticians during a patient's visit to the Neurology Department of Aix-en-Provence Hospital, France. All patients met the UK Parkinson's disease Brain Bank Criteria for diagnosis of idiopathic PD. The average age of the patients at the time of recording was 68.1 ± 7.8 (standard deviation) years; mean PD duration was 8.8 ± 4.7 years.

Patients were studied without (*off*), and then with (*on*) anti-Parkinsonian medication. For this pilot study, we selected the *off* medication condition only. For this condition, patients were recorded after a full overnight fast, *i.e.* at least 12 hours of PD treatment withdrawal. The patients' global motor disability was assessed using the motor section of the Unified Parkinson's Disease Rating Scale (UPDRS, part III, [14]). Mean UPDRS in the *off* medication state was 30.1 ± 10.8. The specific item on speech disability (i.e. UPDRS item 18) was rated on average at 1.5. This item is a 5-point EAI scale (Equal Appearing Interval) where 0 = Normal speech; 1 = Slight loss of expression, diction and/or volume; 2 = Monotone, slurred but understandable; moderately impaired; 3 = Marked impairment, difficult to understand; 4 = Unintelligible. Our results (1.5 *off*) show that the PD speakers are globally slightly dysarthric. The speech of thirty-three control subjects (18 male and 15 female speakers) was recorded in the same manner. Mean age of the control group (C) at the recording time was 67.6 ± 10.8 years.

2.2 *Recordings and task*

Within the Neurology department, a quiet room was set up with EVA2 computerised speech acquisition equipment (SQLab & LPL, Aix-en-Provence, France; www.sqlab.fr/). Speech was recorded with a Headset Cardioid Condenser Microphone with an adjustable headband (AKG C420). Speakers were first instructed to read a standardised text (i.e. "La chèvre de M. Seguin" by A. Daudet) at their most comfortable pitch and loudness, with natural effort. They were then asked to perform a semi-spontaneous task - describing a picture ("Vol de gateau"). All data are part of a larger dysarthric speech corpus collected within the Neurology department [15].

3. Experiment

3.1 Automatic prosodic annotations

In this experiment, we first investigated whether the three Praat scripts that we used for the automatic analysis of speech prosody were suitable and stable enough for Parkinsonian (dysprosodic) speech.

- *Pitch range adaptation script* (De Looze [11]) for the automatic extraction of f0 extrema
- *Syllable nuclei detection script* (De Jong et al. [12]) for the automatic detection of syllable nuclei and measurement of speech rate
- *To TextGrid (silences) Praat command* [13] for the automatic annotation of speech/pause intervals.

3.1.1Pitch range adaptation

Computation of f0 is generally based on default values, in particular in the case of extreme ones. For instance, when creating a Pitch Object in Praat, pitch floor and ceiling parameters are set to the interval [60 – 600 Hz] for general context to [100 – 500 Hz] for female or [75 – 300 Hz] for male voices. Such large pitch ranges, however, generate pitch tracking errors with inappropriate extreme values.

In order to avoid these errors, De Looze [11] proposed a multi-pass process to detect f0. In the first step, the extreme values are set to 60 and 600 Hz. In the second step, the extreme values for a speaker are adapted according to the first pass results: Pitch floor and pitch ceiling are respectively set to the values q15*0.83 (where 'q' stands for percentile) and q65*1.92. These formulae (as well as the formulae q25*0.75 – q75*1.5 and q35*0.72 – q65*1.90) have been shown to provide a better estimation of pitch extrema in 'normal' speech - as they help to exclude more octave errors at the extremes of the f0 distribution - than setting pitch floor and ceiling parameters to larger pitch ranges (e.g. [60 – 600]; [100 – 500] and [75 – 300]). This pitch range adaptation script was therefore used to automatically extract pitch values in our experiment.

3.1.2 Syllable nuclei detection

Measuring articulation rate necessitates annotation in terms of syllables or phonemes. As mentioned in the introduction, this task can however be both time-consuming and error-prone in manual annotations. As a result, several automatic algorithms have been developed for phoneme and syllable detection. Most of them however (e.g. EasyAlign [16]; LIA Aligner [17]) are based on orthographic transcription of speech.

In order to offer a fully automatic process, De Jong et al [12] developed an algorithm that detects syllable nuclei and measures speech rate automatically without requiring any transcription of the speech material as input. In this script, syllable nuclei correspond to peaks in intensity (dB) that are preceded and followed by dips in intensity, with unvoiced peaks being discarded. De Jong's algorithm has been shown to be suitable for the study of non-pathological Dutch speech.

In this experiment, the script was used to automatically annotate syllable nuclei. The script's original output was modified to obtain a TextGrid with a tier annotated in syllables.

3.1.3 Silent pause detection

In order to annotate recordings into interpausal runs, the Praat command *To TextGrid (silences)* [13] was used in this experiment. It provides an annotation of silent and sounding intervals, based on the sound's intensity contour. Several silence thresholds ranging from -20dB to -40dB were tested candidate borders between silence and sounding.

3.2 Automatic prosodic measurements

The prosodic characteristics of Parkinson patients' and of control subjects' speech were investigated in terms of pitch range (level and span), voice intensity (relative level and span) and speech rate (articulation rate and pauses).

Pitch level was measured by calculating the f0-median and given on a linear scale (i.e. Hertz); pitch span was measured by calculating the f0-standard deviation (sd-f0). The intensity of the voice was expressed in terms of the intensity curve median (Int-median) and standard deviation (Int-sd).

Articulation rate was measured by calculating the number of syllable nuclei per second (syllsec). Silent pauses were measured in terms of number and mean pause duration. Number of interpausal runs (IR) and the amount of speech and pause for the recordings were also calculated to further flesh out the description of global speech/pause patterns.

Prosodic features were extracted from two different temporal spans: (1) on the whole recording and (2) for each IR. Differences in f0-median (d_f0-median), sd-f0 (d_sd-f0), Int-median (d_Int-median), Int-sd (d_Int-sd) and syllsec (d_syllsec) between consecutive IR were also calculated.

4. Results

4.1 Evaluation of automatic tools

Pitch range adaptation, Syllable nuclei detection and Silence detection algorithms were first evaluated by comparing manual annotations with automatic annotations. Evaluations were carried out for the 25 control subjects (C) and the 30 Parkinson's Disease (PD) patients in the reading task.

4.1.1Evaluation of pitch range adaptation algorithm

Automatic and manual annotations of pitch extrema were compared using paired t-tests. Results show significant differences between manual and automatic detection of extrema, the mean of the difference being smaller for the pitch range adaptation method than others, i.e. either adjusted to the speaker's gender ([100-500]; [75-300]) or to default values ([60-600]) (cf. *Table 1*). In addition, ANOVA analyses reveal that differences are larger for PD patients than for C subjects (p=0.02).

f0-extrema	Mean of the diff	p-value
minSR	9	3.389e-07
minG	25	1.608e-12
minD	33	5.939e-14
maxSR	-44	6.882e-13
maxG	-82	3.171e-11
maxD	-243	< 2.2e-16

Table 1. Mean of the differences (expressed in Hz) obtained for the minima and maxima extracted with pitch ceiling and floors adjusted to default values (minD; maxD), adjusted according to the speaker's gender (minG; maxG) and adjusted to the speaker's pitch range (minSR; maxSR).

Results show that the proposed pitch range adaptation script [9] provides a better account of f0 measurements than other pitch range settings. However, even with this multi-pass process, differences in Hz are still large (especially for max).

Several f0 extractions are therefore needed to define a speaker's pitch range limits.

4.1.2 Evaluation of syllable nuclei detection

Automatic and manual annotations of syllable nuclei were compared by computing F-measures (F1), calculated as follows:

$$(1) \quad F1 = 2 * \frac{Pr*Re}{Pr+Re}$$

where *Pr* denotes the precision (ratio of hits to all hits and false alarms) and *Re* the recall (ratio of hits to all hits in the set). Hits correspond to syllable nuclei detection within 20ms of the manually annotated syllables while missed values when outside 20ms.

F-measures were significantly smaller for PD patients than for C subjects (p= 0.00776), which means that syllable detection is less accurate for the PD population than it is for C subjects. However, as for both populations F-measures are relatively high (C=0.90; PD=0.86), the annotation in terms of syllable nuclei can be considered reliable.

4.1.3 Evaluation of silent detection

Automatic and manual annotations of speech/pause intervals were compared by computing F-measures and Hamming distances (Hd). F-measures were computed using equation *(1)*, while Hd were found by calculating the distance between two binary vectors (respectively corresponding to manual and automatic annotations) with pauses coded as zeros and speech as ones.

Among the different tested thresholds, results show that for our corpus a threshold set at -30dB provides higher F-measures and smaller hamming distances. In addition, results show that for this threshold, F-measures (F1) are high enough and Hd small enough to provide a reliable automatic annotation in terms of speech/pause intervals (F1=0.76; HD=0.10). In addition, no significant difference for Hd and F-measure were found between PD and C subjects.

4.1.4 Evaluation of extracted prosodic features

Automatic annotations were also evaluated by comparing the values of the prosodic features extracted from manual and automatic annotations using paired t-tests. Significance was set at p<0.01.

With respect to global measurements: T-tests based on global measurements show that articulation rate and mean pause durations obtained with automatic annotation (AA) are not significantly different from those obtained with manual annotation (MA) for both population. Moreover, number of pauses and number of runs are not significantly different in AA and MA for PD patients. Results show however that mean run durations measured from AA are significantly different from those obtained from MA for both populations.

With respect to interpausal runs (IR) measurements: T-tests based on IR show that prosodic values extracted from automatic annotations (AA) are not significantly different from those extracted from manual annotations (MA), except for Int-median and syllsec. Moreover, no significant differences are found between PD and C subjects for all parameters extracted. Prosodic values extracted from AA are thus as reliable for PD patients as they are for control subjects.

4.2 Prosodic characteristics of Parkinsonian speech

In order to investigate the prosodic characteristics of Parkinsonian read and semi-spontaneous speech, the data were analysed using the 'R' software using linear mixed-models, with Type (PD vs. Control) and Task (reading vs. semi-spontaneous) as fixed factors and Speaker as random factor. P-values were estimated using MCMC sampling and significance was considered for p<0.01. Tables 2 and 3 provide the mean values obtained for each prosodic parameter according to Type and Task as well as the absolute differences of these mean values obtained between Parkinson (PD) and Control (C) subjects in read (R) and semi-spontaneous (S) task.

4.2.1Global measurements

Analyses based on global measurements revealed a highly significant main effect of Type on mean pause duration (mpauseD) and speech/pause ratio (sp_ratio). Results show that mpauseD and pause_ratio are higher for PD patients than they are for C subjects. No significant differences between C and PD subjects were found for f0-median, sd-f0, mean run duration, number of pauses and number of runs. Analyses also revealed a highly significant effect of Task on median_f0, mpauseD, pause_ratio and number of pauses. Results show that f0-median and npauses are higher in the R than in the S task; In addition, pause_ratio is smaller for R than for S.

An interaction between Type and Task reached high significance for speakers' mean pause duration and speech/pause ratio. No interaction was found between Type and Task for the other prosodic parameters.

	C/R	C/S	PD/R	PD/S	C-PD/R	C-PD/S
median_f0	155	142	155	147	0	5
sd_f0	25,215	26,826	22,649	23,883	2,566	2,943
syllsec	5,124	4,944	4,864	4,921	0,26	0,023
mpauseD	0,461	0,659	0,522	1,14	0,061	0,481
mrunD	1,11	1,128	1,172	1,129	0,062	0,001
sp_ratio	30,575	38,377	31,799	50,54	**1,224***	**12,163***
npause	45,818	31,843	45,799	26,113	0,019	5,73

*Table 2. Mean values obtained with global measurements for each prosodic parameters according to Type and Task: C/R for control subjects in reading task, C/S in semi-spontaneous task; PD/R for PD patients in reading task and PD/S in semi-spontaneous task; C-PD/R and C-PD/S stand for the absolute differences between Control and PD subjects in read and semi-spontaneous tasks respectively. Significant differences between PD and C subjects are highlighted in bold and followed by *.*

4.2.2 Interpausal runs

Analyses based on interpausal runs (IR) revealed a highly significant effect of Type on differences between consecutive IR, in terms of f0-median, sd-f0 and syllsec. Results show that d_f0-median, d_sd-f0 and d_syllsec between runs are smaller for PD than they are for C. No significant differences were found between C and PD subjects in terms of IR f0-median, sd-f0 and syllsec, neither in terms of difference between consecutive IR in terms of Int-median and Int-sd.

Analyses also revealed a highly significant effect of Task on all prosodic parameters. IR f0-median and syllsec are shown to be higher in reading task (R) than in semi-spontaneous task (S); Differences in f0-median, sd-f0, syllsec, Int-median and Int-sd between IR are lower in R than in S.

An interaction between Type and Task reached high significance for IR f0-median, sd-f0 and syllsec and for consecutive IR differences in f0-median, sd-f0, syllsec, Int-median and Int-sd: f0-median is lower, sd-f0 is narrower, d_f0-median, d_sd-f0 and d_syllsec are smaller in S for PD patients than for C subjects; syllsec, d_Int-median and d_Int-sd for S task are higher for PD patients than for C subjects.

	C/R	C/S	PD/R	PD/S	C-PD/R	C-PD/S
median_f0	155	146	155	148	0,516	**1,32***
d_median_f0	18,437	22,338	13,412	14,703	**5,025***	**7,635***
sd_f0	19,767	20,718	18,375	17,68	1,392	**3,038***
d_sd_f0	7,879	11,168	6,786	8,67	**1,093***	**2,498***
syllsec	5,093	5,032	4,872	5,092	0,221	**0,06***
d_syllsec	1,556	2,032	1,498	1,817	**0,058***	**0,215***
d_Intmedian	3,607	3,857	2,74	4,022	0,867	**0,165***
d_Int-sd	2,466	2,521	2,518	2,438	0,052	**0,083***

*Table 3. Mean values obtained with respect to IR for each prosodic parameters according to Type and Task: C/R for control subjects in reading task, C/S in semi-spontaneous task; PD/R for PD patients in reading task and PD/S in semi-spontaneous task; C-PD/R and C-PD/S stand for the absolute differences between Control and PD subjects in read and semi-spontaneous tasks respectively. Significant differences between PD and C subjects are highlighted in bold and followed by *.*

4. Discussion and conclusion

In this paper, we have presented an automatic analysis of the prosodic variations in the speech of 25 control and 30 Parkinsonian subjects (in *off* (non-medicated) state) in reading and semi-spontaneous tasks. We first evaluated the reliability of existing algorithms for the automatic analysis of speech prosody and have shown that these tools are efficient for the analysis of Parkinsonian speech if a large amount of data is processed and if measurements rely on the extractions of prosodic features at different time intervals. Our evaluation of the Pitch range adaptation script [11] reveals it is all the more important for pitch extraction extrema, for which several candidates at the extreme of the f0-curve must be considered in order to define a speaker's pitch range limits.

In addition, our study reveals that Parkinsonian speech is characterised by longer pauses in both reading and spontaneous tasks, and by lower pitch level, narrower pitch range and faster articulation rate in spontaneous task, for which the difference in terms of pause length between control and PD subjects is all the more important. These findings corroborate those of the literature [1-9] with the exception that we found no difference in terms of pitch range and articulation rate between control and PD speakers in reading task. This could be explained by the fact that PD patients in our experiment are slightly dysarthric, their dysprosodic characteristics being only revealed in semi-spontaneous speech. These findings highlight the importance of evaluating PD patients in different tasks.

Our study also reveals that differences in terms of pitch range and articulation rate between consecutive interpausal runs are smaller in PD patients than in control subjects in both reading and spontaneous speech, these differences between IR being all the more smaller in semi-spontaneous speech. Our results further show that differences in voice intensity level between IR are smaller and in voice intensity range larger in PD speech than in control subjects' in semi-spontaneous task. These findings come to augment the current descriptions of Parkinsonian speech, where the reported monotonous voice is shown to be related not only to a narrowing of the pitch range but also to smaller variations in pitch range. Similarly, our results suggest that the reported monoloudness in PD speech could be due to smaller variations in voice intensity level. One may assume that because prosodic variations are smaller in PD speech, some communicative functions (e.g. topic changes, speakers' attitudes) may not be conveyed in read and in semi-spontaneous Parkinsonian speech.

Our study thus demonstrates that differences in prosodic features between consecutive interpausal runs better capture the prosodic characteristics of Parkinsonian speech than global measurements or interpausal runs' prosodic features. In the literature, most prosodic measurements have been made from entire recordings or from smaller temporal domains. Investigating prosodic variations by calculating the differences between interpausal runs could therefore augment these established measurements, in particular when dealing with slightly dysarthric speech.

Acknowledgements: This research was supported by ANR BLAN08-0125 of the French National Research Agency.

6. References

[1] Canter, G. J. Speech characteristics of patients with Parkinson's disease: Intensity, pitch and duration, *J. Speech Hearing Des*., Vol 28, pp 217-224. 1963.

[2] Monrad-Krohn GH. Dysprosody or altered melody of language. *Brain.* 70: 405-415, 1947.

[3] Darley FL, Aronson AE, Brown JR. Motor speech disorders. W.B. Saunders and Co., 1975.

[4] Weismer, G., Acoustic description of dysarthric speech: Perception correlates and physiological inferences, *Rosenbeck, C. J. (ed), Seminar in speech and language, Thieme Stratton*, New York, p 324. 1984.

[5] Caekebeke, J. F. V., Jenekens-Schinkel, A., van der Linden, M. E., Buruma, O. J. S., & Roos,R. A. C. The interpretation of dysprosody in patients with Parkinson's disease. *Journal of Neurology, Neurosurgery, and Psychiatry,* **54,** 145–148. 1991.

[6] Ho Ak, Iansek & Bradshaw. Regulation of parkinsonian speech volume: the effect of interlocutor distance. *Journal of Neurology & Psychiatry*: 67: 199-202. 1999.

[7] Teston & Viallet. L'évaluation objective de la prosodie in Les dysarthries, Ed : Auzou et al. Problèmes en médecine de rééducation. Masson. Paris p 109-121. 2001.

[8] Duez, D. (2006). Syllable structure, syllable duration and final lengthening in parkinsonian French speech. Journal of Multilingual Communication Disorders, vol. 4, no. 1. 2006, p. 45-57.

[9] Cheang, H.S. and Pell, M.D. An acoustic investigation of Parkinsonian speech in linguistic and emotional contexts, *Journal of Neurolinguistics,* 20, 221-241. 2007.

[10] Kent R., Ball M., Voice quality measurement, Singular Publ. Group, Inc., 492 p. 2000.

[11] De Looze, C. Analyse et interprétation de l'empan temporel des variations prosodiques en français et en anglais contemporain. Doctoral thesis, Université de Provence, 2010.

[12] De Jong, N. H. & Wempe, T. Praat script to detect nuclei and measure speech rate automatically. *Journal of Behavior research methods* : Vol. 41, Nr. 2 pp. 385-390, 2009.

[13] Boersma, P. & Weenink, D. Praat: doing phonetics by computer [Computer program], 2011 http://www.praat.org/.

[14] Fahn, S., Elton, R. L., & Members of the UPDRS Development Committee. Unified Parkinson's Disease Rating Scale. In S. Fahn, C. D. Marsden, D. B. Calne, & M. Goldstein (Eds.), *Recent development in Parkinson's disease*: Vol. 2 (pp. 153- 164). Newark, NJ: Macmillan Healthcare, 1987.

[15] Ghio, A.; Pouchoulin, G.; Teston, B.; Pinto, S.; Fredouille, C.; De Looze, C.; Robert, D.; Viallet, F.; Giovanni, A. How to manage sound, physiological and clinical data of 2500 dysphonic and dysarthric speakers? *Speech Communication*, 16p., 2011.

[16] Goldman, J.-P ., *EasyAligner: a semi-automatic phonetic alignment tool under Praat*, 2007 Available at http://latcui.unige.ch/phonetique.

[17] Audibert,N.;Fougeron,C.;Fredouille,C.;Meunier,C.;Panseri, O. Evaluation d'un alignement automatique sur la parole dysarthrique. *Journées d'Etude sur la Parole*, p1-4 2010.