PROSODIC CHANGES IN THE SPEECH OF A SINGLE SPEAKER WITH PARKINSON'S DISEASE

Vass Verkhodanova[1, 2], Matt Coler[1], Roel Jonkers[3], Kees de Bot[4], Wander Lowie[3, 2]

[1]Campus Fryslân, University of Groningen, The Netherlands

[2]Research School of Behavioural and Cognitive Neurosciences, University of Groningen, The Netherlands

[3]Center for Language and Cognition Groningen, University of Groningen, The Netherlands

[4]University of Pannonia, Hungary

v.verkhodanova@rug.nl, m.coler@rug.nl, r.jonkers@rug.nl, c.l.j.de.bot@rug.nl, w.m.lowie@rug.nl

ABSTRACT

This contribution describes our research into how Parkinson's Disease (PD) impacts the production and perception of speech. We performed a longitudinal study, making a time series of monthly recordings of the same individual with PD over a year. To determine if the change in prosody would be noticeable both on production and perception levels, we performed acoustic analysis of prosodic features and a perceptual experiment with short phrases taken from the recordings as stimuli. The results of the acoustic analysis showed a decline in f_0 variation towards the end of the time period. The results of the perceptual experiment demonstrated that listeners rated the later recordings as less healthy relative to the earlier ones. Listeners' experience with speech disorders influenced the trend, which was more pronounced for the experienced listeners compared to the listeners with no prior experience with speech disorders.

Keywords: Parkinson's Disease, prosody, speech perception, expert knowledge

1. INTRODUCTION

Parkinson's Disease (PD) is a progressive neurological disorder caused by the progressive death of dopaminergic cells in the brain [12]. Among motor manifestations, such as resting tremor, muscle rigidity and bradykinesia, PD patients often develop a speech disorder – hypokinetic dysarthria – resulting from disturbances in muscular control over the speech mechanism [4]. The most studied and described changes second to PD are f_0 deviations commonly referred to as "monopitch" [8, 20, 3], distorted rhythm of speech [21], reduced intensity of voice or "monoloudness" [8, 20, 3], and a hoarse and breathy voice quality [23].

In this study, we explore both the longitudinal effect of PD on speech prosody of a single non-dysarthric speaker and how healthy listeners perceive his speech. To those ends, we performed monthly measurements of the same PD speaker over a year.

There are several longitudinal studies focusing on speech of PD speakers, most of which analysed recordings collected at two time points with intervals between them ranging from seven months [21] to 3.7 years [11]. The results demonstrated reduction in pitch variability [21], instability of steady syllable repetition [19], increased speech rate [11], deteriorations of quality of voice and articulatory velocity and precision [19]. One study described a longitudinal analysis of speech in a single PD speaker over an 11-year period (seven years prior to diagnosis of PD, and three years post-diagnosis) based on archives of national television [9]. Results suggest that changes in f_0 variability can be detected as early as five years prior to diagnosis [9].

We selected the three characteristics which were most indicative of PD speech and allowed easily for automatic measurements: f_0, speech rate, and voice quality. Prosodic changes due to PD has been reported to have similar patterns in different languages [15, 18, 20], with changes in f_0 being most prominent and most studied. Literature on prosody perception in speech affected by PD is scarce and usually involves patients already diagnosed with dysarthria. In two papers concerned with perception of harsh and rough voice of people with PD, this characterisric was reported among the most severely affected dimensions [15, 25], whereas the variable speech rate was not cited among severely affected characteristics. According to the literature, it neither noticeably influenced intelligibility, nor was perceived as affected both in off and on medication states [13, 6, 25].

To determine if longitudinal changes in speech of a single PD speaker without diagnosis of dysarthria would be detectable, we approached the question from two perspectives: one related to acoustics and one related to perception. For the former we hypothesised if any acoustic changes are to manifest themselves within the year, monopitch would be one of them. For the latter, we hypothesised that healthy listeners would be able to perceive a difference in the speaker's voice provided the presence of sufficient acoustic differences along one of the three aforementioned characteristics. Additionally we expected that listeners trained in speech and language pathology would be more sensitive to such changes relative to listeners who lacked this expertise [10].

To test these hypotheses we collected data (2.1), designed a perception experiment (2.2-2.4), and performed an acoustic analysis (2.5).

2. METHODS

2.1. Data collection

Speech recordings were obtained from one male native Dutch speaker who is fluent in English, and who uses both languages daily. At the start of the speech recordings the participant was 66 years old. He was diagnosed with PD six years prior to beginning of the recordings. He has not been diagnosed with hypokinetic dysarthria, but has a history of stuttering.

Recordings consisted of five speech tasks: sustained phonation of the vowel /a/, interview with an open question, picture and video descriptions (one of Heaton pictures and Charlie Chaplin clip), and reading (North Wind and the Sun passage). All tasks were performed first in English and subsequently Dutch. The recordings were collected every month to the extent possible (mean interval is 5.2 weeks, SD = 2.2) from one to three hours after medication intake. The recording sessions took place in quiet rooms at the university with the Zoom H2 recorder placed at around a 40 cm distance. Though perceptual experiment was conducted with both Dutch and English stimuli, we only examined the Dutch data relative to the hypotheses of the current study. The collection and analysis of the material was approved by the Medical Ethics Committee of the University Medical Center Groningen.

2.2. Participants for perceptual experiment

Of thr 61 native Dutch listeners who participated in the experiment, there were people with different experience with speech disorders. Based on their experience and training we divided them into two groups: "naïve" listeners with no prior experience with speech disorders (hereafter the "naïve" group) and speech therapists and/or students of neurolinguistics with experience in listening to disordered speech (hereafter the "experienced" group). The naïve group consisted of 51 people (mean age 27.5, SD 7.7 years). The experienced group consisted of 10 people (mean age 24.4, SD 1.4 years). All participants reported normal hearing.

2.3. Stimuli

We used fragments of 2-3 seconds taken from the Dutch and English spontaneous monologues and reading tasks of five sessions out of 12 (days 0, 107, 204, 286, 411). From each of five sessions we selected six samples: four fragments from spontaneous monologues (two per language) and two fragments from reading tasks (one per language). The total amount of stimuli was 30 phrases selected according to three criteria: 1) they should not include artefacts of stuttering, 2) they should consist of at least four words, 3) they should be extracted from declarative statements. To the extent possible, fragments from monologues were extracted from the first and second half of the recording.

2.4. Procedure

Participants completed a rating task in which they listened to the stimuli in randomised order. Participants were told that they would hear short phrases and were asked to rate them on a 7 Likert scale according to their perception of healthiness (from "very healthy" to "very unhealthy"). The experiment was built within the OpenSesame program [16]. The procedure consisted of a short practice session and the main part. In the practice session, to get participants acquainted with the task they were asked to rate two stimuli of two different voices: one healthy and one affected by dysarthria. For the main part there were 30 stimuli of our PD speaker. The speech samples were intensity normalized and presented over headphones (Koss Pro4S.) Participants could listen to each sample as many times as they wanted.

2.5. Acoustic analyses

To determine whether the acoustic changes are evident within the year on a prosodic level, we performed an acoustic analysis of the selected speech aspects of Dutch monologues and reading. The definition of all analysed acoustic parameters and details of their measurements are summarized in Table 1.

Table 1: Overview of parameters and their measurement methods

Parameter	Description	Method of measurement
f_0 coefficient of variation	Variance of fundamental frequency (f_0), representing the variations of vibration rate of vocal folds	RAPT [22]
Speech rate	The number of syllables per total time	Praat script [5]
Articulation rate	The number of syllables produced per speaking time	Praat script [5]
Jitter	Frequency perturbation, representing the extent of variation of the voice range	Praat [2]
Shimmer	Amplitude perturbation, representing rough speech	Praat [2]
RPDE	Recurrence period density entropy, representing the inefficiency of voice frequency control	Algorithm [14]
HNR	Harmonics-to noise ratio, representing voice hoarseness. HNR is defined as the amount of noise in the speech	Praat [2]

2.5.1. *f_0 estimation*

Pitch tracking was performed with the Talkin's RAPT algorithm [22] implemented in the SPTK toolkit for Python [1]. The RAPT algorithm identifies pitch candidates with the cross-correlation function and then attempts to select the "best fit" at each frame by dynamic programming [17, 22]. From the pitch trajectory we calculated the f_0 coefficient of variation (CV, variance corrected for the means) to estimate the speaker's f_0 range.

2.5.2. *Speech rate*

Measuring speech and articulation rates requires annotation of phonemes or syllables. This procedure is time-consuming and sometimes error-prone. Therefore, we measured speech and articulation rates automatically by detecting syllable nuclei with a Praat script written by de Jong et al. [5]. In this script, syllable nuclei correspond to peaks in intensity preceded and followed by dips in intensity, with unvoiced peaks being discarded. The script has been shown to be informative for the study of French [7] and Dutch [24] dysarthric speech. We have used a -20 dB silence threshold, 4 dB dip and 70 ms as a minimal pause duration. Speech rate was computed as the number of syllables divided by total time, and articulation rate as number of syllables divided by phonation time.

2.5.3. *Voice quality*

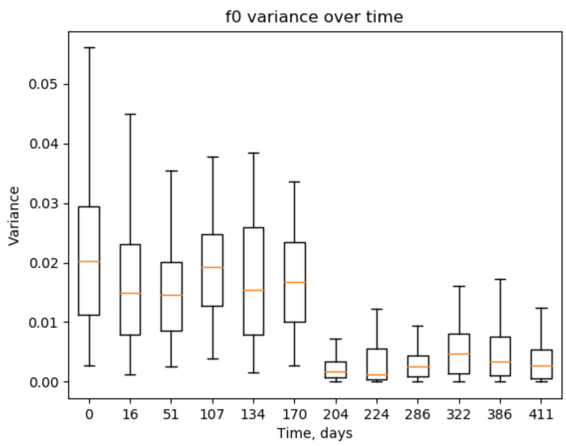
For voice quality we analyzed recordings of the sustained phonation, measuring jitter, shimmer, recurrence period density entropy (RPDE [14]) and harmonics-to-noise ratio (HNR). All of them were measured automatically either with Praat or with the algorithm implemented in Python (see table 1).

3. RESULTS

3.1. Results of acoustic analyses

The analysis of coefficient of variation for f_0 showed a decline from the beginning to the end of the sessions (see Fig.1). A simple linear regression showed that the decline was significant (F = 205.5, $p <$ .001), with R^2 of 0.14 and slope of -4.41×10^{-5}. Speech and articulation rate showed no trends, nor did measurements for RPDE and HNR. Shimmer did not show any significant decline, while jitter did: F = 6.2, $p < 0.03$, with R^2 of 0.18 and slope = -3.58×10^{-5}.

Figure 1: f_0 variance based on CV measurements during the session for all 12 sessions

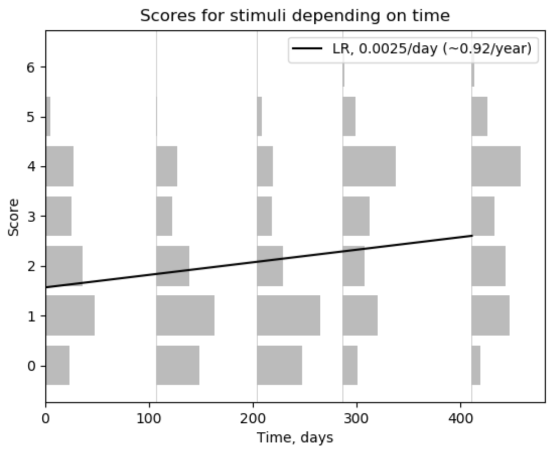


3.2. Results of rating patterns

To assess the rating patterns of the participants we fitted a simple linear regression model in R. A significant regression equation predicting scores depending on time (F = 52.42, $p < .001$) with R^2 of 0.054 and slope coefficient 0.0025. (see Fig.2). To see if there was a difference between naïve and experienced groups we fitted separate linear models. For the naïve group, the regression equation showed significance (F = 36.5, $p < .001$), with R^2 of 0.046; slope coefficient was 0.0023. For the experienced group, the regression equation was significant as well (F = 18.4, $p < .001$), with R^2 of 0.11; the slope coefficient was 0.0039.

Fitting separate models for subsets of stimuli from monologue and reading tasks showed that both groups had steeper slopes for monologues than for reading (0.0026 vs 0.0015 for the naïve group and 0.0041 vs 0.0034 for the experienced group), the model for the naïve group rating stimuli from the reading task did not reach significance ($p > 0.05$).

Figure 2: Dependence of scores for stimuli on time for the healthy listeners



To determine whether the results of the linear regression were not random, we applied a Monte Carlo analysis. In the performed simulation we modelled the probability of different slope outcomes. We randomized the scores 1000 times and calculated the slope for every randomized set of scores. The resulted distribution of slopes had a mean value of 1.3×10^{-5} and SD of 0.0003 with a standard error of 1.14×10^{-5}. We also performed a resampling technique based on the jackknife resampling to evaluate the possibility of bias. We calculated slopes for 1/3 of the data set 1000 times and found that variance for slopes was extremely small: 2.78×10^{-7}.

4. DISCUSSION

In this study we explored the question of longitudinal changes in speech of a single PD speaker without the diagnosis of dysarthria. Acoustic analysis showed no significant changes for speech or articulation rates, shimmer, RPDE or HNR. Significant changes were present in f_0 and jitter, both of which are related to pitch, since jitter is periodicity measurement, relying on f_0, and f_0 is acoustic component of pitch. Our findings are in line with reported role of monopitch being one of the earlier symptoms of dysarthria. A prominent dip in variance of f_0 CVs appeared unexpectedly. We interviewed our PD participant at every session, and he did not reported any (life) events that could have affected his speech within the month before the dip. We have found neither changes in the recording procedure, nor noise conditions that could affect the data. This leads us to the interpretation that possibly some physical change did take place affecting the speech of our participant, that might have triggered the decline. It is too early to say if it could be an onset of dysarthria. Further research into other aspects of speech such as vowel and consonant articulation might shed some light on this hypothesis. General trajectory of the f_0 variation decline is in agreement with our speaker's neurologist's impression of a very slow but steady decline.

The results of the perceptual experiments validated the acoustic analysis, showing the trend of rating later recording as less healthy. Difference in rating between naïve and experienced groups is an interesting finding, that will be addressed in future studies. Trends resulted in linear regression analysis are significant. The R^2 values were expected to be lower because of the nature of the data: the spread of the scores is quite broad while we were looking for the slight changes in the rating patterns. There was no apparent effect of the dip in variance of f_0 CVs on the rating patterns, suggesting that f_0 difference is not prominent enough to guide raters categorisation on its own.

We have found longitudinal changes in speech both by means of acoustic analysis and a perceptual experiment, proving our initial hypotheses on monopitch and perception. Although our research targetted one individual with PD, the results indicate a clear benefit to speech production and prosody tracking in PD speakers, which may help in the early detection of dysarthria in PD.

5. ACKNOWLEDGEMENTS

We thank the students from neurolinguistics master programme at the University of Groningen and all the participants who volunteered to participate in our experiment.

6. REFERENCES

[1] Sptk: The speech signal processing toolkit (version 3.11). http://sp-tk.sourceforge.net/.

[2] Boersma, P., Weenink, D. 2017. Praat, a system for doing phonetics by computer (version 6.0. 28).

[3] Bunton, K., Kent, R. D., Kent, J. F., Duffy, J. R. 2001. The effects of flattening fundamental frequency contours on sentence intelligibility in speakers with dysarthria. *Clinical Linguistics & Phonetics* 15(3), 181–193.

[4] Darley, F. L., Aronson, A. E., Brown, J. R. 1969. Differential diagnostic patterns of dysarthria. *Journal of Speech, Language, and Hearing Research* 12(2), 246–269.

[5] De Jong, N. H., Wempe, T. 2009. Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior research methods* 41(2), 385–390.

[6] De Letter, M., Santens, P., Estercam, I., Van Maele, G., De Bodt, M., Boon, P., Van Borsel, J. 2007. Levodopa-induced modifications of prosody and comprehensibility in advanced Parkinson's disease as perceived by professional listeners. *Clinical Linguistics & Phonetics* 21(10), 783–791.

[7] De Looze, C., Ghio, A., Scherer, S., Pouchoulin, G., Viallet, F. 2012. Automatic analysis of the prosodic variations in Parkinsonian read and semi-spontaneous speech. *Speech Prosody 6th International Conference* 4.

[8] Galaz, Z., Mekyska, J., Mzourek, Z., Smekal, Z., Rektorova, I., Eliasova, I., Kostalova, M., Mrackova, M., Berankova, D. 2016. Prosodic analysis of neutral, stress-modified and rhymed speech in patients with Parkinson's disease. *Computer methods and programs in biomedicine* 127, 301–317.

[9] Harel, B., Cannizzaro, M., Snyder, P. J. 2004. Variability in fundamental frequency during speech in prodromal and incipient Parkinson's disease: A longitudinal case study. *Brain and cognition* 56(1), 24–29.

[10] Harris, R., Leenders, K. L., de Jong, B. M. 2016. Speech dysprosody but no music 'dysprosody' in Parkinson's disease. *Brain and language* 163, 1–9.

[11] Huber, J. E., Darling-White, M. 2017. longitudinal changes in speech breathing in older adults with and without Parkinson's disease. *Seminars in speech and language* volume 38. Thieme Medical Publishers 200–209.

[12] Kalia, L. V., Lang, A. E. 2016. Parkinson's disease in 2015: evolving basic, pathological and clinical concepts in PD. *Nature reviews Neurology* 12(2), 65.

[13] Kuo, C., Tjaden, K., Sussman, J. E. 2014. Acoustic and perceptual correlates of faster-than-habitual speech produced by speakers with Parkinson's disease and multiple sclerosis. *Journal of communication disorders* 52, 156–169.

[14] Little, M. A., McSharry, P. E., Roberts, S. J., Costello, D. A., Moroz, I. M. 2007. Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. *Biomedical engineering online* 6(1), 23.

[15] Ma, J. K.-Y., Whitehill, T. L., So, S. Y.-S. 2010. Intonation contrast in cantonese speakers with hypokinetic dysarthria associated with Parkinson's disease. *Journal of Speech, Language, and Hearing Research* 53(4), 836–849.

[16] Mathôt, S., Schreij, D., Theeuwes, J. 2012. OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior research methods* 44(2), 314–324.

[17] Morrison, D., Wang, R., De Silva, L. C. 2007. Ensemble methods for spoken emotion recognition in call-centres. *Speech communication* 49(2), 98–112.

[18] Rusz, J., Cmejla, R., Ruzickova, H., Ruzicka, E. 2011. Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated parkinson's disease. *The journal of the Acoustical Society of America* 129(1), 350–367.

[19] Skodda, S., Flasskamp, A., Schlegel, U. 2011. Instability of syllable repetition as a marker of disease progression in Parkinson's disease: a longitudinal study. *Movement disorders* 26(1), 59–64.

[20] Skodda, S., Rinsche, H., Schlegel, U. 2009. Progression of dysprosody in parkinson's disease over time: a longitudinal study. *Movement disorders: official journal of the Movement Disorder Society* 24(5), 716–722.

[21] Skodda, S., Schlegel, U. 2008. Speech rate and rhythm in Parkinson's disease. *Movement disorders: official journal of the Movement Disorder Society* 23(7), 985–992.

[22] Talkin, D. 1995. A robust algorithm for pitch tracking (RAPT). *Speech coding and synthesis* 495, 518.

[23] Tsanas, A., Little, M. A., McSharry, P. E., Ramig, L. O. 2010. Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. *IEEE transactions on Biomedical Engineering* 57(4), 884–893.

[24] Verkhodanova, V., Coler, M. 2018. Prosodic and segmental correlates of spontaneous dutch speech in patients with Parkinson's disease: A pilot study. *Speech Prosody 9th International Conference* 163–166.

[25] Whitehill, T. L., Ma, J. K.-Y., Lee, A. S.-Y. 2003. Perceptual characteristics of Cantonese hypokinetic dysarthria. *Clinical linguistics & phonetics* 17(4-5), 265–271.