Acoustic, Phonetic and Prosodic Features of Parkinson's disease Speech

Jorge Proença[1], Arlindo Veiga[1,2], Sara Candeias[1], Fernando Perdigão[1,2]

[1] Instituto de Telecomunicações, Coimbra, Portugal

[2] Electrical and Computer Eng. Department, University of Coimbra, Portugal

{jproenca,aveiga,saracandeias,fp}@co.it.pt

Abstract. *One of the frequent symptoms of Parkinson's disease (PD) is a progressive dysarthria, the difficulty with articulation. Characterization of dysarthric speech is important, for instance, for the adaptation of speech recognition systems to PD patients. In this work, we investigate acoustic and phonetic-prosodic characteristics of speech produced by PD patients in the European Portuguese. For this purpose, a speech database, consisting mainly on phonetically rich sentences, has been collected from a group of patients with different degrees of PD severity and a healthy control group. Only vowels in a stressed position in continuous speech context (i.e. no sustained vowels) were analyzed. Results show a centralization of vowel formant frequencies for PD speech, as expected. However, some of the usual features for discriminating PD speech were not always found to be statistically significant. Furthermore, an analysis of acoustic, spectral and prosodic features towards classifying PD speech shows that dynamic features are of highest importance in this task.*

1. Introduction

Parkinson is a type of degenerative disease of the central nervous system and is more common in the elderly, with most cases occurring after the age of fifty [Fahn 2010]. Some of the most obvious symptoms are related to movement impairment, including *bradykinesia* (or slowness of movement), resting tremor, rigidity and difficulty on walking [Lees et al. 2009]. Typically 90% of patients with PD reveal disabilities in speech production [Ramig et al. 2008]. The most common speech problems experienced involve *hypophonia* (or reduced volume), monotone (or reduced pitch range) and dysarthria (difficulty with articulation of sounds and syllables) [Ramig et al. 2008; Goberman and Coelho 2002]. Among the effects of the PD, the patients also show difficulty to handle common computer peripherals (keyboard, mouse, touchscreen, etc.). As a result of the reduced speech intelligibility and movement rigidity, people with PD often become isolated [Cote et al. 2000] and speech technologies can offer a relevant contribution to improve their quality of life.

The present study aims to identify some acoustic, phonetic and prosodic characteristics of the speech of PD patients for European Portuguese. The ultimate goal is to adapt a speech recognizer for PD patients to be used for friendly technological interfaces.

Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology, pages 205–209,
Fortaleza, CE, Brazil, October 21–23, 2013.

In order to reveal possible correlations among vowel articulation and the stage of the disease, [Skoda et al. 2011] analyzed both F1 and F2 frequency values of the vowels /a/, /i/ and /u/, representing the triangle vertices of the vowel chart. Based on the values of the F1 and F2 of these three vowels, geometric calculations such as vowel space area (VSA) [Skoda et al. 2011; Roy et al. 2009] and vowel articulation index (VAI) [Skoda et al. 2011] have been used as indicators of a dysarthric speech. Among the differences between PD and healthy speech related to the first two formants, differences in the patterns of variability for vowel production were marked in [Soares 2011]. The extension of these findings for European Portuguese language awaits clarification so far. Considering the specific speech characteristics typically associated in the literature with PD speech, we focus here on the detailed spectral parameters of the speech produced by PD patients. Phonetic information relies on F1 and F2 for the vowels [i], [E], [a], [O] and [u] (in SAMPA, [Wells 1997]). Finally, several acoustical measures were studied in order to identify features that are well correlated with PD

In the next section, the speech corpus of PD patients collected for experimental findings is described. Section 3 describes the phonetic analysis on the corpus. . In section 4 acoustic and prosodic features are analyzed for classification of PD speech. The paper ends with the concluding remarks.

2. Parkinson Disease Speech Corpus

A corpus consisting of a series of 1002 utterances of phonetically rich sentences and isolated words (mainly commands) was collected from 22 PD patients (12 females and 10 males) with different degrees of the disease. Patients are aged between 50 and 80 years old. Recordings took place at the neurology service of the Hospital of the University of Coimbra. In each recording sessions, the same common set-up was used, which consists of a laptop computer and three microphones. Parkinson disease speech corpus represents 90 minutes of read speech recorded at a 48 kHz sampling rate. A healthy control group (3 females and 4 males between 25 and 51 years old) was also recruited for recording the same battery of speech production tasks under identical acoustic conditions. Segmentation at the phone-level was automatically done for each session through forced-alignment with our phone recognizer [Lopes et al. 2008].

It is possible to distinguish two different PD speech degrees on the basis on the fluency of the speech produced: normally articulated and rhythmical (Low-PD), and slow and awkwardly articulated speech (High-PD). Hence, the patients were distinguished as belonging to one of these two levels of dysarthria. This classification was done by perceptual experiments with a high level of agreement.

3. Phonetic Analysis

3.1. Vowel Formants

Given the characteristics of continuous speech, estimating formant frequency values is not a straightforward task. To minimize problems with unclearly outlined or rapidly variable formants, restrictions were made on which segments would be proper for analysis. Therefore, vowels [a], [E], [i], [O] and [u] in a stressed position were selected from the aligned transcriptions at phone level, including only those with duration above

50ms. Furthermore, each segment was cut where the energy level was 20 dB below of the maximum energy, to specifically consider the well established part of the vowel.

The Praat tool [Boersma and Weenink 2013] was used to automatically extract the first (F1) and second (F2) formant frequencies. A similar method to [Escudero et al. 2009] was applied, given the foreknowledge that different vowels and speakers need different formant ceilings for the automatic calculation. An iterative calculation of formants was performed in 10 ms steps using ceilings in the 4000-5500Hz range (for males) or 4500-6500Hz (for females) in 50Hz steps. The optimal ceiling for a vowel of a given speaker was the one that provided the minimum variance of F1 and F2 between the available samples of that vowel. This was calculated as the sum of the variances of 20log(F1) and 20log(F2), where F1 and F2 are the median values of each segment. Through manual verification, it was deemed necessary to further restrict the ceiling ranges of the method, depending on vowel, ranges decided empirically: (Female/Male) 5500-6000/4800-5200Hz for [i] and [E], 4800-5200/4300-4700Hz for [a] and 4000-4500/4000-4200 for [O] and [u]. After calculations in 10 ms steps, we chose to only extract the median formant values for each vowel. The number of vowel tokens kept after restrictions was 4555 ([i]: 1233; [E]: 666; [a]: 1941; [O]: 376; [u]: 339).

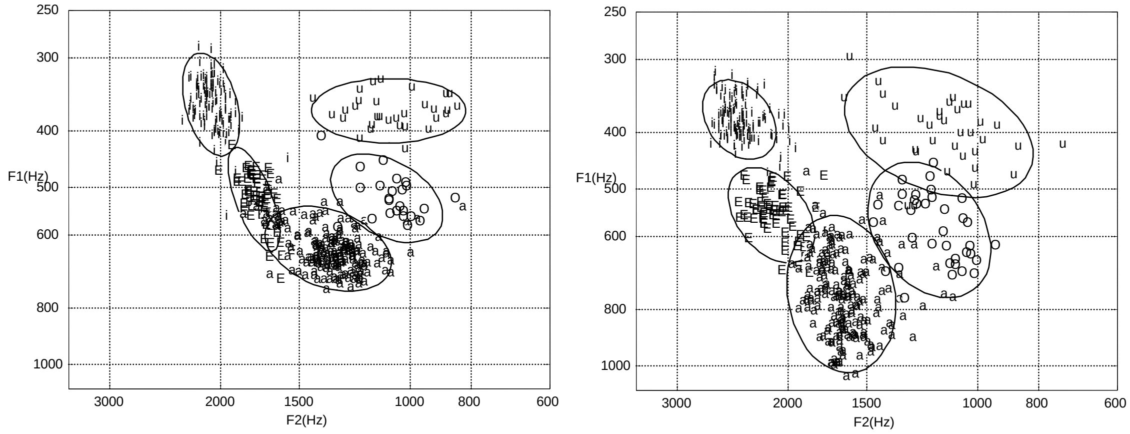

Figure 1: F1 and F2 for vowels [i], [E], [a], [O] and [u] of Males of the Low-PD group (left) and Females of the Low-PD group (right).

F1 and F2 values for the vowels show large variations but don’t overlap much. Figure 1 shows an example of these values for the Low-PD group. Some of the [u] F2 values are uncharacteristically high, and many of these cases were manually confirmed to be correct. This can be explained by the circumstance of continuous speech, where the vowels may be partially displaced to a centralized position. Median values were calculated as an effective way to remove some outliers. Comparing the F1/F2 vowel triangles for High-PD and Low-PD speakers (Fig. 2), it can be reported that as the dysarthria progresses the triangle of the vowels reduces, with lower F1 values, mainly for [E], [a] and [O], and with a slight centralization of F2 values. This tendency could confirm the articulatory restriction of PD patients as a result of the rigidity of the vocal muscle, considerably evident in the production of open central vowel [a]. Male PD speakers also presented a larger centralization of [O] and [u]. In general, the articulation of vowels of the low-PD group tends to be closer to the control group, indicating that the reduction of the vowels space cannot be considered a feasible indicator for mild forms of dysarthria.

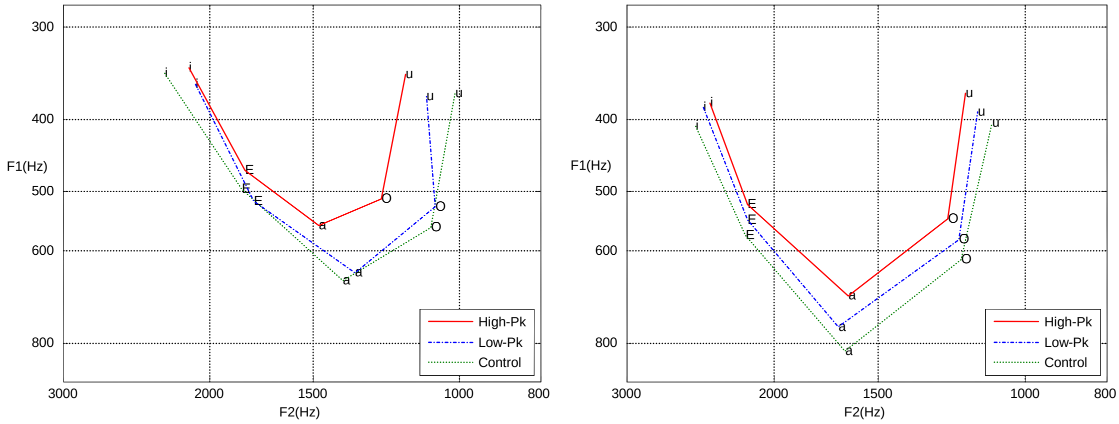

Figure 2: F2 and F1 median values of [i], [E], [a], [O] and [u] for males (left) and females (right) in Control, Low-PD and High-PD groups.

3.2 Formant metrics

For each speaker, vowel space area (VSA) and vowel articulation index (VAI) were calculated, [Skoda et al. 2011]. Another unused metric is the Formant Centralization Ratio (FCR), which is the inverse value of VAI. Although average values show an expected reduction of vowel space area and lower articulation for PD-patients, the only result of statistical significance is Male VAI for Control (0.92±0.07) vs. High-PD (0.80±0.07) with p=0.038.

4. Acoustic-Prosodic analysis

To identify the acoustical characteristics which may be appropriate to classify speech from PD patients, two sets of features were considered: one is based on Gaussian Mixture Models (GMM) supervectors applied to Mel Frequency Cepstral Coefficients; the other is derived from the openSMILE toolkit [Eyben et al. 2010] which corresponds to measures applied to spectral and prosodic parameters. To evaluate a figure of merit for each feature in terms of PD disease discriminative performance, we considered a two class problem trained to separate speakers with high-PD level from others. Two classifiers were used: SVM for GMM supervectors and Bayes Network, [Hall et al. 2009], for the “Smile” features. Summarily, the main indication from the results was that dynamic features were the highest ranked in separating PD speech from normal.

5. Conclusions

In this paper we described our current research towards finding acoustic and phonetic characteristics in the speech of Parkinson disease (PD) patients. A speech database from different levels of severity of PD has been collected for this purpose. F1 and F2 formant frequencies have shown differences in central vowels for high level of PD; however, from the usual metrics to evaluate dysarthric speech, only vowel articulation for control males versus high-level PD males had statistical significance. We also found that the most significant features for PD speech discrimination are dynamic features. This is in accordance with the effect of movement impairment in PD, which reduces the capability of articulating in a dynamic way. We plan to evaluate the possibility of integrating these features on the development of a speech recognizer system adapted to PD patients.

References

Boersma, P. and Weenink, D., “Praat: doing phonetics by computer” Computer program, Version 5.3.42, http://www.praat.org/, retrieved on 2 March 2013.

Cote L., Sprinzeles LL., Elliott R. and Kutscher AH [Eds], Parkinson’s Disease and Quality of Life. New York: Haworth Press, 2000.

Escudero, P., Boersma, P., Rauber, A. S. and Bion, R. A. H., “A cross-dialect acoustic description of vowels: Brazilian and European Portuguese”, Journal of the Acoustical Society of America, 126(3): 1379-1393, 2009.

Eyben, F., Wöllmer, M. and Schuller, B., “openSMILE - The Munich Versatile and Fast Open-Source Audio Feature Extractor,” ACM Multimedia Proc.: 1459–1462, 2010.

Fahn, S., “Parkinson's disease: 10 years of progress, 1997–2007”, Mov Disord, 25 (Suppl 1): S2–S14, 2010.

Goberman, A. and Coelho, C., “Acoustic analysis of Parkinsonian speech I: Speech characteristics and L-Dopa therapy”, NeuroRehabilitation, 17: 237–246, 2002.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I., “The WEKA Data Mining Software: an Update” SIGKDD Explorations, 11, 2009.

Lees, A. J., Hardy, J. and Revesz, T., “Parkinson's disease”, Lancet, 373: 2055–2066, 2009.

Lopes, J., Neves, C., Veiga, A., Maciel, A. M., Lopes, C., Perdigão, F., Sá, L. V., "Development of a Speech Recognizer with the Tecnovoz Database", Proc International Conf. on Computational Processing of Portuguese - PROPOR, Aveiro, Portugal, Vol. , pp. 260 - 263, September, 2008.

Ramig LO., Fox C., and Sapir S., “Speech treatment for Parkinson's disease”, Expert Rev Neurother, 8(2):297-309, 2008.

Roy, N., Nissen, S.L., Dromey, C. and Sapir, S., “Articulatory changes in muscle tension dysphonia: Evidence of vowel space expansion following manual circumlaryngeal therapy”, Journal of Communication Disorders, 42(2): 124–135, 2009.

Skodda, S., Visser, W. and Schlegel, U., “Vowel Articulation in Parkinson's Disease”, Journal of Voice, 25(4): 467-472, 2011.

Soares, M. F. de Paula, “Vowel variability in speakers with parkinson’s disease”, ICPhS XVII Proc.: 1570-1573, 2011.

Wells, J.C., 1997. 'SAMPA computer readable phonetic alphabet'. In Gibbon, D., Moore, R. and Winski, R. (eds.), 1997. *Handbook of Standards and Resources for Spoken Language Systems*. Berlin and New York: Mouton de Gruyter. Part IV, section B