2010 International Symposium on Nonlinear Theory and its Applications
NOLTA2010, Krakow, Poland, September 5-8, 2010

New nonlinear markers and insights into speech signal degradation for effective tracking of Parkinson's disease symptom severity

Athanasios Tsanas[1,2], Max A. Little[1,2,3], Patrick E. McSharry[1,2,4] and Lorraine O. Ramig[5,6]

[1] Systems Analysis, Modelling and Prediction (SAMP) group, Mathematical Institute and Department of Engineering Science, University of Oxford, Oxford, UK, [2]Oxford Centre for Industrial and Applied Mathematics (OCIAM), Mathematical Institute, University of Oxford, Oxford, UK, [3]Oxford Centre for Integrative Systems Biology, Department of Physics, University of Oxford, Oxford, UK, [4]Smith School of Enterprise and the Environment, University of Oxford, UK, [5]Speech, Language, and Hearing Science, University of Colorado, Boulder, Colorado, USA, [6]National Center for Voice and Speech, Denver, Colorado, USA

Email: (A. Tsanas) tsanas@maths.ox.ac.uk, tsanasthanasis@gmail.com, (M. Little) littlem@maths.ox.ac.uk, (P. McSharry) patrick@mcsharry.net, (L. Ramig) Lorraine.Ramig@colorado.edu

Abstract– We have recently shown that speech signal degradation can be used to quantitatively predict average Parkinson's disease (PD) symptom severity, which is typically evaluated on the Unified Parkinson's Disease Rating Scale (UPDRS). In this study, we demonstrate the potential of wavelets to reveal changes in fundamental frequency variations with PD progression. We develop a set of new measures based on wavelets, energy, and entropy, which form robust indicators of the UPDRS. These results demonstrate that PD leads to dissimilar speech patterns in males and females, tentatively taken to indicate different patho-physiological mechanisms.

1. Introduction

Parkinson's disease (PD) is a neurodegenerative disorder affecting approximately 100 people for every 100,000 in the population [1]. Since the probability for PD onset increases steeply over the age 50 [2] and given that the population worldwide is growing older, these estimates could increase further in the near future. Although PD is progressive and ultimately fatal, pharmacological and surgical treatments are available to alleviate some of the symptoms and slow down disease progression. Therefore, to optimize treatment, early diagnosis and frequent PD progression tracking are essential [3]. PD symptoms include tremor, rigidity and general deterioration of muscle control. The diagnosis of *Parkinsonism* is given when these symptoms can be attributed to neurotoxins or drugs; when the aetiology is unknown, the disease is termed *idiopathic*.

At present, PD symptoms are physically assessed by clinical experts using empirical tests and guidelines. The clinical rater's assessment is typically expressed using the gold standard metric *Unified Parkinson's Disease Rating Scale* (UPDRS) [4]. For untreated patients the UPDRS spans the numerical range 0-176, with 0 representing symptom-free (healthy person) and 176 total disability. The UPDRS consists of three *components*: (1) Mentation, Behavior and Mood; (2) Activities of daily living; and (3) Motor (muscular control). Collectively, the three components are known as *total UPDRS*. The third component is known as *motor UPDRS*, and ranges from 0-108, with 0 indicating no motor symptoms and 108 denoting total lack of motor control. The UPDRS assessment by clinical experts is costly to national health systems relying on human expertise, and often cumbersome to patients who have to physically visit the clinic over regular intervals. These factors impose the need for accurate, objective, remote tracking of average PD symptom severity.

Degraded speech performance has been qualitatively related with PD at least since the beginning of the 1960s [3], however, only recently strong evidence has emerged linking *speech degradation* with *PD progression* [5], [6]. Those studies prompted us to investigate the statistical mapping of a range of classical speech signal processing algorithms (known as *dysphonia measures* in the jargon of the speech literature) to UPDRS [7], [8]. In this study, we explore the effectiveness of wavelets to reveal changes in fundamental frequency variations with PD progression.

2. Methods

2.1. Data

We used data from the original study of Goetz *et al.*, in which 52 subjects with idiopathic PD diagnosis within the past five years were recruited to a clinical trial [6]. Subjects were given a PD diagnosis if they had at least

two of the following symptoms: rest tremor, bradykinesia or rigidity, with no evidence of Parkinsonism. Their symptom severity was expressed using the UPDRS at three intervals: baseline, three-months and six-months into the trial. The recruited subjects were followed for six months during which they were asked to complete a series of tests *weekly*, using the Intel At Home Testing Device (AHTD). Among these tests, the subjects were required to sustain the vowel 'ahh …' for as long as possible and as steadily as possible. Four sustained vowel phonations were recorded at a level of loudness that was comfortable for the subject and two at twice the level of loudness that was comfortable on each day the recruits took the test. We used data from subjects that had completed at least 20 valid study sessions. Table 1 summarizes the details of the 42 PD subjects used in this study. After initial screening to remove flawed phonations (too short, patient coughing), we processed 5,875 signals using dysphonia measure signal processing algorithms implemented in the Matlab software package.

Table 1: Summary of the AHTD data

	MALES (28 subjects)	FEMALES (14 subjects)
Age (years)	64.8 ± 8.1	63.6 ± 11.6
Weeks since PD diagnosis	63.0 ± 61.9	89.7 ± 81.2
Motor-UPDRS (baseline, 3-months, 6-months)	20.3 ± 8.5 21.9 ± 8.7 22.0 ± 9.2	17.6 ± 7.4 21.2 ± 10.5 20.1 ± 9.4
Total-UPDRS (baseline, 3-months, 6-months)	27.5 ± 11.6 30.4 ± 11.8 31.0 ± 12.4	24.2 ± 9.1 27.4 ± 12.1 26.8 ± 10.8

Figures are given in the form mean ± standard deviation.

2.2. Discrete wavelet transform and wavelet decomposition

Wavelets have the property of quantifying regularity effects (scale aspects) and transient processes (time aspects), qualities which make them well suited for detecting scale and time deviations from an expected pattern. The rationale is that a healthy person is expected to be able to sustain a vowel with minimal deviation from exact periodicity, whilst people with pathological voices cannot [9]. Moreover, wavelet decomposition is well adapted to the study of fractal properties and self-similarity of signals, properties of speech signals used previously in developing dysphonia measures [10].

The discrete wavelet transform (DWT) expresses the initial signal using *approximation* and *detail* coefficients. The *wavelet decomposition* then is successive expressing the approximation coefficients using subsequent layers to extract new *approximation* and *detail* signals. The layers of the wavelet decomposition are known as *levels*. Practically speaking, the resulting wavelet coefficients can be thought of as similarity (resemblance) indices between the selected wavelet and the signal in each level, where large coefficients denote large resemblance. For more details regarding wavelets we refer to [11].

2.3. Extracting features based on wavelets

As a first pre-processing step, we extracted the fundamental frequency F_0 from each of the 5,875 signals. Algorithms extracting F_0 focus on a time window of the original signal (the window can be either pre-specified, or it can be determined by the F_0 algorithm, e.g. using zero-crossing). Then, for each of those windows, the F_0 is estimated giving an F_0 series vector. There are many algorithms to compute F_0 and this is in itself a topic of intense research [12]; in this study, we used the robust RAPT algorithm [13]. Then, the input signal vector for wavelet processing is the F_0 series.

We applied wavelet decomposition of the F_0 series in 10 levels, and extracted the wavelet coefficients experimenting with three wavelet families (Daubechies, Symlets, Coiflets). Then, we computed the energy, entropy (using both Shannon's and the log energy definitions), and the Teager-Kaiser Energy Operator (TKEO) for the wavelet coefficients at all levels. The TKEO can be thought of as a nonlinear measure of energy, taking into account both the amplitude and the frequency of the input signal (in this case the wavelet coefficients). It was first proposed in [14] and is defined as:

$$\Psi(x_n) = x_n^2 - x_{n+1} \cdot x_{n-1} \qquad (1)$$

where n denotes the index of the input vector. Then, the TKEO vector gives rise to two features using its mean and its standard deviation for each level. Recently, Little *et al*. have shown that the transformation of the fundamental frequency into its logarithmic perceptual semitone can enhance robustness to confounding factors such as smooth vibrato prior to further processing [15]. Therefore, in addition to the features extracted using the raw F_0series, we computed the log transform of the F_0series and then followed the methodology already outlined to obtain additional features.

The results of all the algorithms using both the raw F_0 series and the log transformed F_0 series are appended in a feature vector, which is used to characterize each phonation. Essentially, the feature vector reduces the initial vector space with elements equal to the length of the F_0 series to a reduced space, where each element of the feature vector space can be thought of as a distinctive *feature*. This process was repeated for all the 5,875 phonations where each phonation was characterised by 180 features, resulting in a 5,875×180 *design matrix*.

2.4. Statistical mapping

The UPDRS values in the AHTD study were obtained at baseline, three-month and six-months, whilst the voice recordings were obtained at weekly intervals; therefore we

used a straightforward piecewise linear interpolation to obtain weekly motor-UPDRS and total-UPDRS scores. In doing so, we assumed that the UPDRS did not fluctuate wildly within the three-month intervals between the actual assessments by clinical experts. We have argued in [7], [8] that linear UPDRS progression is the physiologically most plausible disease course *on average*, an assertion supported by other recent studies on early PD [16], [17].

Using the interpolated UPDRS scores, we have the classical supervised learning problem where we want to develop a learner maximizing the accuracy of predicting the response variable y (UPDRS) given the features X (wavelet vector). We have used Breiman's Random Forests (RF) [18], a powerful, nonlinear, non-parametric learner for the statistical mapping of features to UPDRS.

2.5. Feature selection

The use of a large number of features can potentially lead to poor population of the feature space occluding the detection of relevant patterns useful to predict the response variable. This well known problem is known as the *curse of dimensionality*, and is typically addressed by either transforming the initial feature space M into a new space m (where $m < M$), or by selecting K of the initial features in the M-dimensional feature space (m or K are determined by trial and error, e.g. using cross-validation). One of the advantages RF have over alternative learners, is that they rank features internally as an integral part of the statistical mapping process, effectively acting as feature selection *wrappers*. Therefore, we used the ranked sequence of features to decide on the most parsimonious model with performance within one standard error from the optimal (as defined in section 2.6).

2.6. Model validation (cross-validation)

The generalization performance of the proposed model was assessed using 10-fold cross validation with 100 runs. In each run, we randomly split the N phonations (4,010 for males and 1,865 for females). The training set comprises $0.9 \cdot N$ phonations and the testing set comprises $0.1 \cdot N$ phonations. We assess the performance of the learning scheme using the *mean absolute error* (MAE):

$$MAE = \frac{1}{L}\sum_{i \in Q} \left| \widehat{U}_i - U_i \right| \qquad (2)$$

where U_i is the true UPDRS value, $\widehat{U}_i$ is the predicted value, and L is the number of phonations in the dataset denoted by Q, containing the indices of the particular set in each cross-validation run. The MAE over all cross-validation runs was averaged.

3. Results

Having extracted the F_0 series vector, we have experimented using different wavelet families in the DWT step. We found that the wavelets from the Daubechies, Coiflet and Symlet families had similar performance. The out of sample MAE results are summarized in Table 2, and were obtained using the dysphonia measure subsets of Table 3 and the Symlet 4 wavelet.

Table 2: Out-of-sample mean absolute error (MAE) results

Motor UPDRS males	*Total UPDRS males*	*Motor UPDRS females*	*Total UPDRS females*
5.36 ± 0.34	6.93 ± 0.36	4.72 ± 0.38	5.47 ± 0.51

Figures are given in the form mean ± standard deviation.

Table 3: Feature subsets selected using the Random Forests' internal feature ranking property

Males	*Females*
Log entropy 3rd detail coef.	Log entropy 3rd detail coef.
Log entropy 2nd detail coef.	Log entropy 2nd detail coef.
Log entropy 1st detail coef.	Shannon entropy 1st approximation coef. (with prior F_0 transform)
Log entropy 4th detail coef.	Shannon entropy 3rd approximation coef. (with prior F_0 transform)
Shannon entropy 2nd approximation coef.	Log entropy 1st detail coef.
Shannon entropy 1st approximation coef.	Log entropy 4th detail coef.
Log entropy 5th detail coef.	Shannon entropy 2nd approximation coef. (with prior F_0 transform)
Shannon entropy 3rd approximation coef.	Mean TKEO 4th detail coef. (with prior F_0 transform)
Shannon entropy 4th approximation coef.	Energy 4th detail coef. (with prior F_0 transform)
Log entropy 3rd detail coef. (with prior F_0 transform)	Shannon entropy 2nd approximation coef.

4. Discussion

We have introduced a number of new measures based on the use of wavelets to investigate how speech performance degradation can be mapped to UPDRS. We focused on the analysis of the fundamental frequency because this is the single most important characteristic of speech [9], and our previous exploration of the AHTD database [7], [8] has confirmed studies' reporting that F_0 is adversely affected in PD. Our findings demonstrate that the new measures enable replicating the clinical raters' assessments to within 7 UPDRS points for males and within 5.5 UPDRS points for females. This is notable improvement over previous results in [7] and [8] where we used broadly accepted speech signal processing algorithms.

The notable difference observed between the UPDRS estimation performance in males and females suggests that patterns associated with F_0 may be more indicative of PD in female subjects, an argument supported by other F_0 related measures in the AHTD database. Given that higher fundamental frequencies tend to have lower perturbations [19], and that women have on average higher F_0 [9], it is plausible that failure to sustain F_0 periodicity indicates

pronounced voice pathology in females whilst similar deviation from F_0 periodicity could be (at least partly) due to normal vibrato in males. Furthermore, studying Table 3 we can tentatively deduce interesting insights on the most useful PD patterns for both genders. One particularly interesting characteristic is that many of the features in the selected subset for females stem from prior log-transformation of the F_0.

The results of this study support the argument that features extracted based on wavelets are competitive alternatives to the classical dysphonia measures which are currently widely used to analyse dysphonic speech signals.

Acknowledgments

We are grateful to Ralph Gregory for medical insight and to Mike Deisher, Bill DeLeeuw and Sangita Sharma at Intel Corporation, and Paul Moore for fruitful discussions and comments on early drafts of the paper. We also want to thank James McNames, Lucia M. Blasucci, Eric Dishman, Rodger Elble, Christopher G. Goetz, Andy S. Grove, Mark Hallett, Peter H. Kraus, Ken Kubota, John Nutt, Terence Sanger, Kapil D. Sethi, Ejaz A. Shamim, Helen Bronte-Stewart, Jennifer Spielman, Barr C. Taylor, David Wolff, and Allan D. Wu, who were responsible for the design and construction of the AHTD device and organizing the trials in which the data used in this study was collected.

Declaration

A. Tsanas is funded in part by Intel Corporation and in part by the Engineering and Physical Sciences Research Council (EPSRC).

References

[1] S. von Campenhausen, B. Bornschein, R. Wick, K. Bötzel, C. Sampaio, W. Poewe, W. Oertel, U. Siebert, K. Berger, and R. Dodel, "Prevalence and incidence of Parkinson's disease in Europe," *European Neuropsychopharmacology*, Vol. 15, pp. 473-490, 2005

[2] A. Elbaz, J. H. Bower, D. M. Maraganore, S. K. McDonnell, B. J. Peterson, J. E. Ahlskog, D. J. Schaid, and W. A. Rocca, "Risk tables for parkinsonism and Parkinson's disease," *Journal of Clinical Epidemiology*, Vol. 55, pp. 25-31, 2002

[3] R. Pahwa, R., and E. Lyons E. (eds.), *Handbook of Parkinson's Disease*, 4th edition, Informa Healthcare, USA, 2007

[4] C. Ramaker, J. Marinus, A. M. Stiggelbout, and B. J. van Hilten, "Systematic evaluation of rating scales for impairment and disability in Parkinson's disease," *Movement Disorders*, Vol. 17, pp. 867-876, 2002

[5] S. Skodda, H. Rinsche, U. Schlegel: "Progression of dysprosody in Parkinson's disease over time – A longitudinal study," *Movement Disorders*, Vol. 24 (5), pp. 716-722, 2009

[6] C.G. Goetz, G.T. Stebbins, D. Wolff, W. Deleeuw, H. Bronte-Stewart, R. Elble, M. Hallett, J. Nutt, L. Ramig, T. Sanger, A.D. Wu, P.H. Kraus, L.M. Blasucci, E.A. Shamim, K.D. Sethi, J. Spielman, K. Kubota, A.S. Grove, E. Dishman, C.B Taylor, "Testing objective measures of motor impairment in early Parkinson's disease: Feasibility study of an at-home testing device," *Movement Disorders*, Vol. 24 (4), pp. 551-556, 2009

[7] A. Tsanas, M. A. Little, P. E. McSharry, L. O. Ramig, "Accurate telemonitoring of Parkinson's disease progression using non-invasive speech tests," *IEEE Transactions Biomedical Engineering*, Vol. 57, pp. 884-893, 2010

[8] A. Tsanas, M. A. Little, P.E. McSharry, L. O. Ramig, "Enhanced classical dysphonia measures and sparse regression for telemonitoring of Parkinson's disease progression," *International Conference on Acoustics, Speech and Signal Processing (ICASSP '10)*, Dallas, Texas, US, pp. 594-597, 14-19 March 2010

[9] I. R. Titze, *Principles of Voice Production*, National Center for Voice and Speech, Iowa City, US, 2nd printing, 2000

[10] M. A. Little, P. E. McSharry, S. J. Roberts, D. Costello, I. M. Moroz, "Exploiting Nonlinear Recurrence and Fractal Scaling Properties for Voice Disorder Detection," *Biomedical Engineering Online* 6:23, 2007

[11] S. Mallat, *A wavelet tour of signal processing*, Academic press, 3rd edition, 2009

[12] R. M. Roark, "Frequency and Voice: perspectives in the time domain," *Journal of Voice*, Vol. 20, pp. 325-354, 2006

[13] D. Talkin, "A Robust Algorithm for Pitch Tracking (RAPT)," in *Speech Coding and Synthesis*, (editor W.B. Kleijn), K.K. Paliwal eds., Elsevier, 1995

[14] J. Kaiser, "On a simple algorithm to calculate the 'energy' of a signal," *International Conference on Acoustics, Speech and Signal Processing (ICASSP '90)*, pp. 381-384, Albuquerque, USA, April 1990

[15] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, L. O. Ramig, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," *IEEE Transactions Biomedical Engineering*, Vol. 56(4), pp. 1015-1022, 2009

[16] W. Maetzler, I. Liepelt, D. Berg, "Progression of Parkinson's disease in the clinical phase: potential markers," *Lancet Neurology*, Vol. 8, pp. 1158-1171, 2009

[17] M. Schüpbach, J. C. Corvol, V. Czernecki, M. B. Djebara, J. L. Golmard, Y. Agid, and A. Hartmann, "Segmental progression of early untreated Parkinson disease: a novel approach to clinical rating," *Journal of Neurology, Neurosurgery and Psychiatry*, Vol. 81, pp. 20-25, 2010

[18] L. Breiman, "Random forests," *Machine learning*, Vol. 45, pp. 5-32, 2001

[19] R. J. Baken and R. F. Orlikoff, *Clinical measurement of speech and voice*, San Diego, Singular Thomson Learning, 2nd ed., 2000