

# ENHANCED CLASSICAL DYSPHONIA MEASURES AND SPARSE REGRESSION FOR TELEMONITORING OF PARKINSON'S DISEASE PROGRESSION

Athanasios Tsanas<sup>1,2</sup>, Max A. Little<sup>1,2,3</sup>, Patrick E. McSharry<sup>1,2</sup>, Lorraine O. Ramig<sup>4,5</sup>

<sup>1</sup>Systems Analysis, Modelling and Prediction (SAMP), Mathematical Institute, University of Oxford, Oxford, UK, <sup>2</sup>Oxford Centre for Industrial and Applied Mathematics (OCIAM), University of Oxford, Oxford, UK, <sup>3</sup>Oxford Centre for Integrative Systems Biology, Department of Physics, University of Oxford, UK, <sup>4</sup>Speech, Language, and Hearing Science, University of Colorado, Boulder, Colorado, USA, <sup>5</sup>National Center for Voice and Speech, Denver, Colorado, USA

## ABSTRACT

Dysphonia measures are signal processing algorithms that offer an objective method for characterizing voice disorders from recorded speech signals. In this paper, we study disordered voices of people with Parkinson's disease (PD). Here, we demonstrate that a simple logarithmic transformation of these dysphonia measures can significantly enhance their potential for identifying subtle changes in PD symptoms. The superiority of the log-transformed measures is reflected in feature selection results using Bayesian Least Absolute Shrinkage and Selection Operator (LASSO) linear regression. We demonstrate the effectiveness of this enhancement in the emerging application of automated characterization of PD symptom progression from voice signals, rated on the Unified Parkinson's Disease Rating Scale (UPDRS), the gold standard clinical metric for PD. Using least squares regression, we show that UPDRS can be accurately predicted to within six points of the clinicians' observations.

**Index Terms**—telemedicine, sparse regression, dysphonia measures, Parkinson's Disease (PD), Least Absolute Shrinkage and Selection Operator (LASSO)

## 1. INTRODUCTION

Parkinson's Disease (PD) is the second most common neurodegenerative disorder after Alzheimer's [1], and it is estimated that more than one million people in North America alone are affected [2]. It is reported that 20/100,000 new cases every year are diagnosed [3] with the incidence rates rising steeply over the age of 50 [4]. At present, there is no available cure, although drugs can alleviate some of the symptoms and slow down the progression of the disease. The classical symptoms of PD are tremor, rigidity and general deterioration of muscle control. Parkinsonism displays the same constellation of symptoms, but these can be attributed to neurotoxins or may be secondary to drugs etc. It has long been known that speech is also affected in people with Parkinson's (PWP); recently, strong evidence has emerged linking *speech degradation* with *PD progression* [5].

PD symptom progression monitoring is typically achieved using a number of empirical tests and physical examinations, which are usually mapped to the gold-standard clinical metric, the

Unified Parkinson's Disease Rating Scale (UPDRS) [6]. In this study, we use the term *motor-UPDRS* to refer to the motor section of the UPDRS (part III), which spans 0-108 with 0 denoting healthy and 108 no motor ability. *Total-UPDRS* refers to the full range of the metric for untreated patients, 0-176, with 176 denoting total disability. The current clinical approach of physical examinations to determine UPDRS is time consuming for clinical staff and awkward for the patients, who have to visit the clinic every three to six months. It is also costly to national health systems.

Speech signal processing algorithms offer an *objective*, potentially reliable method for assessing general voice disorders. In the context of PD, speech signals have been used to separate PWP and healthy controls (people with no PD symptoms) [7], [8], with very encouraging results. We extend these findings to map PD dysphonias to UPDRS scores, an approach suggested by the recent literature [5], [9]. Specifically, we use a range of common, classical measures [10], [11] to extract clinically useful *features*, which objectively characterize the speech signal. We also introduce, as additional features, the logarithmically transformed classical measures. We demonstrate that the log-transformed classical measures convey superior clinical information and are selected by an automatic feature selection algorithm as being more appropriate for UPDRS prediction, outperforming their traditional, non-transformed versions. These findings encourage the adoption of log-transformed classical dysphonia measures in PD symptom telemonitoring.

## 2. METHODS

### 2.1. Data

We used sustained vowel speech recordings from the study of Goetz *et al.*, in which 52 subjects with idiopathic PD diagnosis within the past five years were recruited to a clinical trial [9]. 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. It was deemed safe for the recruits to remain without PD medication for six months. They were physically assessed and given UPDRS scores at baseline, three-months and six-months into the trial. During the six months in which they participated, they were asked to take a series of tests *at home* and at *weekly* intervals, which were directly recorded on the

Intel At Home Testing Device (AHTD), a telemonitoring device. Among these tests, the subjects were required to sustain the vowel ‘ahh ...’ for as long, and as steadily as possible. During each test session, four of these sustained vowel phonations were recorded at comfortable loudness and two at twice the initial loudness. Data from 10 subjects was removed because of insufficient data: the 42 subjects we use in this work had completed at least 20 valid study sessions. After initial screening to remove flawed phonations (too short, patient coughing etc.), we processed 5,875 signals using dysphonia measure signal processing algorithms implemented in the Matlab software package. The 42 subjects (28 males) had an age range (mean  $\pm$  std)  $64.4 \pm 9.24$  years, average motor-UPDRS  $20.84 \pm 8.82$  points and average total UPDRS  $28.44 \pm 11.52$  points.

## 2.2. Feature extraction

The aim of this work is to extract clinically useful information from the sustained vowel phonations, and map them to UPDRS. Collectively, the results of the dysphonia measures for each phonation form a feature vector which is then used as input in a regression setting.

We applied a range of dysphonia measures which have been successfully used in similar problems aimed at separating healthy controls and PWP [8]. We used the widely cited program Praat [10] to compute the classical dysphonia measures, which include quantifying fundamental frequency perturbations (*jitter*), amplitude perturbations (*shimmer*), and signal to noise ratios (*harmonics to noise ratio*). We used the ‘MDVP’ prefix to associate the measures which are equivalent to the results of the Kay Pentax Multi-Dimensional Voice Program. All measures are summarized in Table 1.

The novelty of this study is that we introduce, as additional features, the classical dysphonia measures following logarithmic transformation. The log-transformation of the classical dysphonia measures was suggested by visual inspection of their estimated probability density functions, which invited some form of normalization because these density functions were highly skewed and positive-only. That is, significant details of the dysphonia measures were crowded into a narrow range, and a convenient way of expanding this range while preserving its full diversity is to use logarithmic transformation. Informally, the idea is to force the density functions to become more normally distributed, so that they are more stable input features for the regression scheme we are using to predict UPDRS. The calculation of the measures is exactly the same as with the classical measures, but is followed by log-transforming the results. Although the logarithmic scale has been shown to be more appropriate than the linear scale in some related contexts for physiological reasons (e.g. for voice pitch [9]), our approach here is dictated by purely numerical and statistical considerations.

Fundamentally, each measure tries to extract distinct characteristics of the speech signal, which ideally are non-overlapping. In practice, these measures will be correlated [9] – a concept we will address in section 2.4. The application of each dysphonia measure algorithm produces a single number, giving rise to a  $5,875 \times 26$  matrix.

**TABLE 1: Brief description of the dysphonia measures**

<i>Measure</i>	<i>Description</i>
MDVP: Jitter (%)	Fundamental frequency perturbation (%)
MDVP: Jitter (Abs)	Fundamental frequency perturbation (absolute)
MDVP: RAP	Relative Amplitude Perturbation
MDVP: PPQ	Five-point Period Perturbation Quotient
Jitter: DDP	Average absolute difference of differences between cycles, divided by the average period
MDVP: Shimmer	Local amplitude perturbation
MDVP: Shimmer (dB)	Local amplitude perturbation (decibels)
Shimmer: APQ3	Three point Amplitude Perturbation Quotient
Shimmer: APQ5	Five point Amplitude Perturbation Quotient
MDVP: APQ11	11-point Amplitude Perturbation Quotient
Shimmer: DDA	Average absolute difference between consecutive differences between the amplitudes of consecutive periods
NHR	Noise-to-Harmonics Ratio
HNR	Harmonics-to-Noise Ratio

## 2.3. Linear regression

The recorded UPDRS values were obtained at baseline, three-month and six-months into the AHTD trial, and the voice recordings were obtained at weekly intervals; therefore we need to obtain weekly UPDRS estimates to associate with each phonation. The simplest approach is to use nearest neighbor interpolated UPDRS values, which would imply a sudden jump mid-way between assessments, and physiologically does not seem very plausible. We assume that the UPDRS did not fluctuate wildly within the three-month intervals between physical assessments, and propose a straightforward piecewise linear interpolation to obtain weekly motor-UPDRS and total-UPDRS scores. The interpolation goes exactly through each recorded UPDRS value.

Using the input feature vectors we develop a predictor which maximizes the accuracy of predicting the interpolated UPDRS scores. For simplicity, we choose ordinary least squares regression to map the feature vector  $\mathbf{x} = (x_1, \dots, x_M)$  (where  $M$  is the number of measures) to the UPDRS output  $y$ . We aim to determine the coefficients  $\mathbf{b}$  which minimize the mean squared error between the actual and the predicted UPDRS values:

$$\hat{\mathbf{b}} = \arg \min_{\mathbf{b}} \sum_{i=1}^N \left( y_i - \sum_{j=1}^M x_{ij} b_j \right)^2 \quad (1)$$

where  $N$  is the number of samples (5,875)

## 2.4. Feature selection using the LASSO

A frequent problem in regression and classification settings such as in this study, is the *curse of dimensionality*, where the measures cannot efficiently populate the  $M$ -dimensional feature space given the available  $N$  feature vectors. Thus, a smaller number of input features could potentially provide more accurate results, in

addition to be computationally less expensive. A common manifestation of the curse of dimensionality in regression problems is *collinearity*, where two measures are assigned similar magnitude – opposite signed coefficients assigned to two measures, which effectively cancel each other out. The *Least Absolute Shrinkage and Selection Operator (LASSO)*, proposed by Tibshirani [12], has emerged as a powerful feature selection method [13] based on penalizing the absolute sum of the regression coefficients:

$$\hat{\mathbf{b}}_{Lasso} = \arg \min_{\mathbf{b}} \sum_{i=1}^N \left( y_i - \sum_{j=1}^M x_{ij} b_j \right)^2 + \lambda \sum_{j=1}^M |b_j| \quad (2)$$

where  $\lambda$  is the LASSO *regularization parameter*.

LASSO elects the best subset of input features for the given regularization parameter  $\lambda$  by assigning regression coefficients  $\mathbf{b}$ , some of which will be set to almost exactly zero as a consequence of the sparsity-enhancing property of the absolute sum (and thus effectively removing features). Increasing  $\lambda$  causes additional coefficients to shrink towards zero, further reducing the number of selected features. The optimal value of  $\lambda$  is then typically determined using a grid-search and cross-validation scheme.

### 2.5. Cross-validation

In order to objectively assess the generalization performance of the proposed model we used 10-fold cross validation with 100 runs. In each run, we randomly split the 5,875 phonations: the training set comprises 5,287 phonations which are used to derive the regression coefficients, and the testing set comprises the remaining 588 phonations. We report the error using the *mean absolute error (MAE)*:

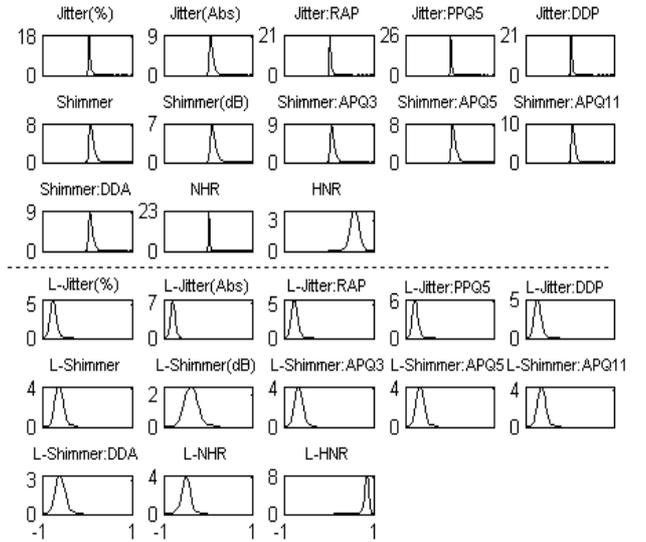
$$MAE = \frac{1}{N} \sum_{i \in Q} |U_i - \hat{U}_i| \quad (3)$$

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

## 3. RESULTS

Fig. 1 displays the probability densities of all the 26 dysphonia measures used in this study (comprising the 13 classical measures and the 13 classical measures after log-transformation). As can be seen, the log-transformed measures are, in general, more normally distributed.

Table 2 presents the out-of-sample MAE results for the 13 classical measures in their traditional form (no log-transformation) and after log-transformation, demonstrating the superiority of the log-transformation approach. It also provides the MAE for the LASSO optimal subset, when the algorithm had to choose the best out of the 26 measures. The selected measures appear in the first column of Table 3. We note that most of the surviving measures are log-transformed.



**Fig. 1. Probability densities of the normalized dysphonia measures (spanning the interval [-1, 1]) for classical dysphonia measures and logarithmically transformed classical measures. Above the dashed line are the original dysphonia measures, and below, in the same relative position, the log-transformed versions. The vertical axes are the probability densities of the normalized measures, estimated using kernel density estimation with Gaussian kernels.**

**TABLE 2: Out-of-sample Mean Absolute Error (MAE) results, with  $\pm$  denoting the standard deviation of the quoted error.**

Dysphonia measures	Motor-UPDRS regression MAE	Total-UPDRS regression MAE
13 classical measures	7.03 $\pm$ 0.17	8.98 $\pm$ 0.25
13 log-transformed measures	6.62 $\pm$ 0.15	8.43 $\pm$ 0.24
LASSO ( $\lambda = 100$ ) optimal set	6.57 $\pm$ 0.16	8.38 $\pm$ 0.23

Finally, Table 3 presents, for the optimal LASSO selected subset typical linear regression coefficients with their standard deviations. The coefficients associated with the log-transformed terms are, in absolute terms, larger than the coefficients of the classical unscaled measures, indicating they are more important in the regression setting.

## 4. SUMMARY AND DISCUSSION

In this study, we introduced the concept of logarithmically transforming classical dysphonia measures, and subsequently mapping them to average PD symptom progression measured using the UPDRS. We have seen that using ordinary least squares regression, the estimated motor-UPDRS was within 6.6 points (out of 108) and total-UPDRS was within 8.4 points (out of 176) of the clinicians' estimates. These are particularly encouraging results, given that all signals were collected in the PWP's homes using self-administered tests.

**TABLE 3: Regression coefficients applying (1) using the optimal subset of measures selected by (2), with  $\pm$  denoting the standard deviation of the quoted coefficients over the 100 cross validation runs.**

Measure	Motor-UPDRS	Total-UPDRS
Shimmer(dB)	$-2.84 \pm 0.24$	$0.61 \pm 0.34$
HNR	$-0.64 \pm 0.02$	$-1.09 \pm 0.03$
Log-Jitter(%)	$4.86 \pm 0.61$	$2.46 \pm 0.82$
Log-Jitter(Abs)	$-11.07 \pm 0.25$	$-5.53 \pm 0.30$
Log-Jitter:PPQ5	$5.48 \pm 0.54$	$2.03 \pm 0.68$
Log-Shimmer:APQ5	$-21.16 \pm 0.47$	$-23.17 \pm 0.65$
Log-Shimmer:APQ11	$24.53 \pm 0.46$	$24.94 \pm 0.61$
Log-NHR	$1.96 \pm 0.20$	$1.79 \pm 0.26$
Log-HNR	$13.44 \pm 0.52$	$31.19 \pm 0.73$

Our aim was to quantify the improvement in UPDRS prediction performance solely due to log-transformation of classical dysphonia measures, but there have been recent developments in non-classical dysphonia measures that capture additional essential speech characteristics [7], [8] and the distributions of these non-classical measures are close to normal, thus, log-transformation of these measures is unnecessary. Nonlinear regression methods could further decrease the estimated error.

In order to reduce the number of measures used as features and determine whether the log-transformed classical measures outperformed the non-transformed measures, we used LASSO regression. This is a sophisticated simultaneous regression and feature selection algorithm that has been shown to be equivalent to brute force search through all possible combinations of measures to minimize the prediction error if there are actually only a few important features [14]. LASSO regression here clearly shows that log transformed classical dysphonia measures convey superior clinical information compared to the raw measures.

## 5. ACKNOWLEDGEMENTS

We are grateful to Mike Deisher, Bill DeLeeuw and Sangita Sharma at Intel Corporation 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.

## 6. REFERENCES

[1] M.C. de Rijk, L.J., Launer, K. Berger, M.M. Breteler, J.F. Dartigues, M. Baldereschi, L. Fratiglioni, A. Lobo, J. Martinez-Lage, C. Trenkwalder, A. Hofman: "Prevalence of Parkinson's disease in Europe: a collaborative study of population-based cohorts", *Neurology*, Vol. 54, pp. 21–23, 2000

[2] A.E. Lang, A.M. Lozano: "Parkinson's disease – First of two parts", *New England Journal Medicine*, Vol. 339, pp. 1044-1053, 1998

[3] M. Rajput, A. Rajput, A.H. Rajput: "Epidemiology" (chapter 2). In *Handbook of Parkinson's disease*, edited by R. Pahwa & K.E. Lyons, 4th edition, Informa Healthcare, USA, 2007

[4] A. Elbaz, J.H. Bowerb, D.M. Maraganoreb, S.K. McDonnell, B.J. Petersona, J.E. Ahlskogb, D.J. Schaidaa and W.A. Rocca: "Risk tables for parkinsonism and Parkinson's disease", *Journal of Clinical Epidemiology*, Vol. 55, pp. 25-31, 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. 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

[7] 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

[8] 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

[9] 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, and C.B: "Testing objective measures of motor impairment in early Parkinson's disease: Feasibility study of an at-home testing device", *Movement Disorders*, Vol. 24 (4), 551-556, 2009

[10] P. Boersma and D. Weenink: Praat: doing phonetics by computer (Version 5.1.15) [Computer program]. Retrieved from <http://www.praat.org/>, 2009

[11] KayPENTAX, Kay Elemetrics Disordered Voice Database, Model 4337, Kay Elemetrics, Lincoln Park, NJ, USA, 1996-2005

[12] R. Tibshirani: "Regression Shrinkage and Selection via the Lasso", *Journal Royal Statistical Society (Series B)*, Vol. 58 (1), pp. 267-288, 1996

[13] T. Hastie, R. Tibshirani, J. Friedman: *The elements of statistical learning: data mining, inference, and prediction*, Springer, 2<sup>nd</sup> edition, 2009

[14] D. Donoho: "For most large underdetermined systems of equations, the minimal L1-norm near-solution approximates the sparsest near-solution", *Communications on Pure and Applied Mathematics*, Vol. 59 (7), 904-934, 2006