1. In your own language, briefly describe, possibly why we can use voice features to detect Parkinson's disease – you may use published peer-reviewed literature using appropriate citations.

According to the paper, there are researches that has shown that around 90% of the people suffering from Parkinson’s disease exhibit some form of vocal impairment [1][2]. So, the paper has considered vocal impairment as a significant indicator of the illness [3]. Also, the measurement of voice is noninvasive and simple to administer. Therefore, voice evaluation can be of significant importance to detect and track the progression of symptoms of Parkinson’s disease.

People suffering from Parkinson’s disease usually display a typical pattern of vocal symptoms like being unable to produce vocal sounds (dysphonia) or speech articulation problem (dysarthria). These voice related disorders have a generic symptom like reduced loudness, breathiness, roughness, decreased energy in the higher parts of the harmonic spectrum and exaggerated vocal tremor.

Citation:

[1] A. K. Ho, R. Iansek, C. Marigliani, J. L. Bradshaw, and S. Gates, “Speech impairment in a large sample of patients with Parkinson’s disease,” *Behav. Neurol.*, vol. 11, pp. 131–137, 1998

[2] J.A.Logemann,H.B.Fisher,B.Boshes,andE.R.Blonsky,“Frequency and co-occurrence of vocal-tract dysfunctions in speech of a large sample of Parkinson patients,” *J. Speech. Hear. Disord.*, vol. 43, pp. 47–57, 1978.

[3] J. R. Duffy, *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management*, 2nd ed. St. Louis, MO: Elsevier, 2005.

2. From the relevant literature (with appropriate citations), describe in your own language, how the features were extracted or, generated from the collected human voice.

The paper presents the estimation measures, traditional and non-standard, for detecting Parkinson’s disease (PD) in a person by detecting dysphonia (voice disorder). A new measure of dysphonia, pitch period entropy (PPE), was introduced in this paper which is robust to many uncontrollable confounding effects including noisy acoustic environments and normal, healthy variations in voice frequency.

For Traditional measures:

The features needed were calculated from the voice signals. The traditional estimation measure used a software called Praat to record the features. The traditional measures based on the application of the short-time autocorrelation to successive segments of the signal, with peak picking to determine the frequency of vibration of the vocal folds (F0 or pitch period), and location in time of the beginning of each cycle of vibration of the vocal folds (pitch marks).

The sequence of frequencies for each vocal cycle are used to derive the jitter and period perturbation by taking the absolute value of the difference between frequencies of each cycle and taking an average of the varying number of cycles. Optionally, the overall average can be used to normalize the data. Shimmer and amplitude are derived from the sequence of maximum extent of the signal amplitude within each vocal cycle. The deviation between cycle amplitudes is given by the average difference of the same sequence. The noise-to harmonic (and harmonics-to-noise) ratios are calculated from the signal-to-noise estimates from the autocorrelation of each cycle.

The second half of each voice signal is dominated by spurious dysphonia (caused mainly by lack of lung pressure), therefore, they are discarding the end part of the phonation [1][2][3].

The authors of this paper are deliberately restricted to relative (or perturbative) measures of pitch period and amplitude since they are more robust to uncontrollable environmental and individual variations.

For Non-traditional Measures:

The correlation dimension (D2) is calculated by first time-delay embedding the signal to recreate the phase space of the nonlinear dynamical system that is proposed to generate the speech signal [4].

DFA is a measure of the extent of the stochastic self-similarity of the noise in the speech signal. The DFA algorithm calculates the extent of amplitude variation F (L) of the speech signal over a range of time scales L, and the self- similarity of the speech signal is quantified by the slope α of a straight line on a log–log plot of L versus F (L). A simple nonlinear transformation then normalizes these slope values (αn o r m ) to the range [0, 1] [5].

For new measure of dysphonia, pitch period entropy (PPE):

The measurements of abnormal speech pitch variation need to take into account the two important effects: healthy, smooth vibrato and microtremor, and the logarithmic nature of speech pitch in speech production (and perception).

To implement these two insights algorithmically, the authors first obtain the pitch sequence of the phonation and converted to the logarithmic semitone scale p(t), where p is the semitone pitch at time t. Then they analyzed the roughness of variations in this sequence over and above any healthy, smooth variations, by first removing linear temporal correlations in this semitone sequence with a standard linear whitening filter (coefficients of which are estimated using linear prediction by the covariance method [7]) to produce the relative semitone variation sequence r(t).

Subsequently, they constructed a discrete probability distribution of occurrence of relative semitone variations P (r). And finally, they calculated the entropy of this probability distribution [8], which then characterizes the extent of (non-Gaussian) fluctuations in the sequence of relative semitone pitch period variations.

An increase in this entropy measure better reflects the variations over and above natural healthy variations in pitch observed in healthy speech production.

Citation:

* [1] P. Boersma and D. Weenink, “Praat, a system for doing phonetics by computer,” *Glot Int.*, vol. 5, pp. 341–345, 2001.
* [2] KayPENTAX,“Kayelemetricsdisorderedvoicedatabase,model4337,” Kay Elemetrics, Lincoln Park, NJ, 1996–2005.
* [3] P.Boersma,“Accurateshort-termanalysisofthefundamentalfrequency and the harmonics-to-noise ratio of a sampled sound,” presented at the Inst. Phonet. Sci., University of Amsterdam, Amsterdam, The Netherlands, 1993, vol. 17.
* [4] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, New ed. Cambridge, U.K.: Cambridge Univ. Press, 1999.
* [5] M.A.Little,P.E.McSharry,S.J.Roberts,D.A.Costello,andI.M.Moroz, “Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection,” *Biomed. Eng. Online.*, vol. 6, p. 23, 2007.
* [6] B. C. J. Moore, *An Introduction to the Psychology of Hearing*, 5th ed. Boston, MA: Academic, 2003.
* [7] J.G.ProakisandD.G.Manolakis,*DigitalSignalProcessing:Principles, Algorithms, and Applications*, 3rd ed. Upper Saddle River, NJ: Prentice- Hall, 1996.
* [8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ: Wiley, 2006.

The main traditional measurement methods include F0 (the fundamental frequency or pitch of vocal oscillation), absolute sound pressure level (indicating the relative loudness of speech), jitter (the extent of variation in speech F0 from vocal cycle to vocal cycle), shimmer (the extent of variation in speech amplitude from cycle to cycle), and noise-to-harmonics ratios (the amplitude of noise relative to tonal components in the speech)