Efficiency of Voice Features Based on Consonant for Detection of Parkinson's Disease

Conference Paper · October 2018

DOI: 10.1109/LSC.2018.8572266

9 authors, including:

Rekha Viswanathan
RMIT University

Parham Khojasteh
RMIT University

Behzad Aliahmad
RMIT University

Sridhar P Arjunan
SRM Institute of Science and Technology

Efficiency of Voice Features based on Consonant for Detection of Parkinson's Disease

R. Viswanathan*[1], P. Khojasteh[1], B. Aliahmad[1], S.P. Arjunan[1], S. Ragnav[2], P. Kempster[2], D. K. Kumar[1]
[1] Biosignals Lab, School of Engineering, RMIT University, Melbourne, Australia
[2] Monash Medical Centre, Clayton, Melbourne, Australia
*Corresponding Author Email: s3427376@student.rmit.edu.au

***Abstract*- The objective of the study is to determine the efficiency of features extracted from sustained voiced consonant /m/ in the diagnosis of Parkinson's Disease (PD). The diagnostics applicability of the phonation /m/ is also compared with that of sustained phonation /a/, the one which is commonly employed in PD speech studies. The study included 40 subjects out of which 18 were PD and 22 were controls. The features extracted were used in SVM classifier model to differentiate PD and healthy subjects. The phonation /m/ yielded classification accuracy of 93% and Matthews Correlation Coefficient (MCC) of 0.85 while the classification accuracy for phonation /a/ was 70% and MCC of 0.39. The spearman correlation coefficient analysis also showed that the features from /m/ phonation were highly correlated with the Unified Parkinson's Disease Rating Scale (UPDRS-III) motor score. The results suggest the applicability of features corresponding to nasal consonant in the diagnosis and progression monitoring of PD.**

Keywords—Parkinson's Disease; Voice features; Consonant; Vowel.

I. INTRODUCTION

Parkinson's disease (PD) is a progressive neurological condition generally affecting aged population [1]. The most widely used tool in determining the symptoms and severity of PD is the Unified Parkinson's disease scale (UPDRS) which requires clinical expertise and experience. Assessment based on UPDRS is subjective, time-consuming. It also requires the patients' physical presence in the clinic which makes it unsuitable for remote regions. Thus there is a need for an objective, reliable and easy to use diagnostic modality [2].

Motor speech disorder is seen developing in 90% of people with PD (PWP) which is termed as Hypokinetic Dysarthria. This affects multiple levels of speech including phonation, articulation, prosody, and intelligibility. Hypokinetic Dysarthria affects laryngeal function, articulatory and respiratory muscles. Most of the time, PWP will not be aware of the issues with their communication. The voice of PD patients is associated with low volume, monotone, high pitch, tremulous, breathiness, hoarseness, and sometimes inability to spontaneously maintain loudness [3-5].

Majority of the studies for the diagnosis of PD have used sustained phonation /a/ as the speech task and a few studies have even included /e/, /i/, /o/, and /u/ phonation. The features extracted from the sustained phonation were either conventional (which includes jitter, shimmer, harmonics to noise ratio, measure of formant bandwidth, mean formant frequencies etc.) or non-conventional features (which includes empirical mode decomposition, correlation dimension, Hurst exponent etc.) or a combination of them[6-8]. One of the major limitations in using sustained vowel phonation /a/ is that it cannot provide much information about the articulators [9]. From the previous studies, it is shown that the voiced consonants are prone to higher distortion in PD and the subjects exhibited articulatory difficulties in consonant articulation; thus making imprecision of consonants as one of the major problems in PWP[10, 11].

The production of consonants especially voiced nasal consonants requires co-ordination of timing of breathing, larynx, articulators and contact force between lips, tongue, and palate[12].Thus the analysis of nasal consonants provides joint analysis of temporal co-ordination, timing, and laryngeal control. The consonants like /b/, /d/, /t/ etc. have been studied in terms of voice onset time (VOT) in PWP to determine their variation between PD and control groups and it has found that there is difference in VOT between PWP and control subjects [13, 14].

In this work, we have considered consonants as it has been tested and reported in the previous studies [13,14]. The main objective of the study is to determine the ability of dysphonia features extracted from voiced nasal consonants in differentiating PWP and control subjects. The study uses sustained nasal consonant /m/ (as in ham) together with sustained phoneme /a/ (as in car). In this study we have performed the following a) compare the classification ability of voice features extracted from sustained nasal consonant /m/ and vowel /a/ in distinguishing PWP and control subjects b) evaluate the strength of association of the features from both the phonation with UPDRS III score using Correlation analysis.

II. MATERIALS AND METHODS

A. *Study protocol*

The ethics for this study was approved by The Monash Health Human Research Ethics Committee (HREC), Monash

Medical Centre, Clayton, Victoria, Australia. (LNR/16/MonH/319) and performed in accordance with Declaration of Helsinki.

The study included collection of sustained vowel phonation /a/ and sustained phonation /m/ from control subjects and PWP who are diagnosed based on Queen Brain Bank Criterion (QBBC) for clinical assistance in the assessment of PD.

The study participants were seated comfortably in a room for data collection. A smart phone and a wired head worn omni-directional microphone were used to collect the data. The voice samples were recorded in .wav format with a sampling rate of 48kHZ and 16-bit resolution.

Informed written consent was obtained prior to the recording along with participant demographics and clinical information. The PWP were all in off-state and between 1-10 years of PD duration without any other major psychiatric and active medical condition. The control group did not also have any psychiatric illness and active medical condition. UPDRS-III score was evaluated for all the study participants. For the phoneme task, the subjects were advised to take a deep breath before the phoneme and sustain the phoneme as much as they can. A rest period of 60 seconds was provided between each phoneme task.

B. Participants

A total of 40 subjects which included 18 PWP (9 males, 9 females; Age range: 57-83) and 22 control subjects (14 males, 8 females; Age range: 56-84) participated in the study. The control subjects were aged matched with the PWP group. The speech task included recording of sustained vowel /a/ and sustained consonant /m/ following the study protocol. An example of sustained /a/ and /m/ waveform for both PD and control subject is illustrated below in Fig. 1 and Fig. 2.

C. Feature reduction and classification

The database contained single sustained /a/ and /m/ phonation from each study subject. Prior to feature extraction, the unvoiced segments from the start and end of the signals were removed. Over 60 features were extracted from the phonemes. The majority of features extracted were based on [2] and in addition to that, 9 glottal source signal features were also included in the study. Table 1. shows the overall features extracted from /a/ and /m/.

To overcome the curse of dimensionality, Relief-F feature reduction was applied to the feature set of /a/ and /m/ phonation. Top ranked 10 features for /a/ and /m/ respectively were then selected and fed to the classification model. SVM classifier with Radial Basis Function (RBF) kernel was used to differentiate PWP and control subjects based on the reduced feature set. Leave one out validation method was employed to evaluate the general performance of the classification model. Matthews Correlation coefficient (MCC) was also calculated to measure the quality of the binary classifier used in the study.

Anderson-Darling normality test was used to determine the distribution of the features. The strength of association between the reduced features and UPDRS-III score was evaluated using Spearman correlation coefficient test.

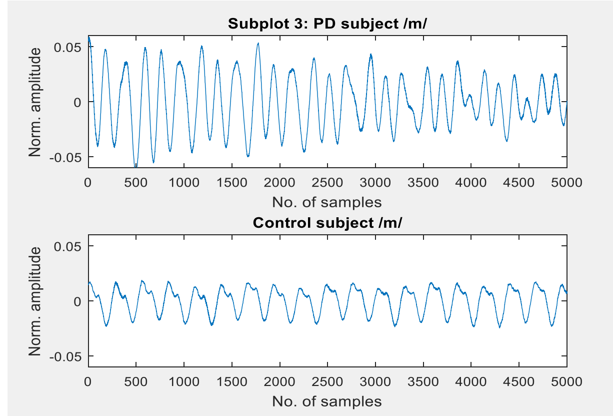

Fig.1. Sustained /m/ for PD and control subject

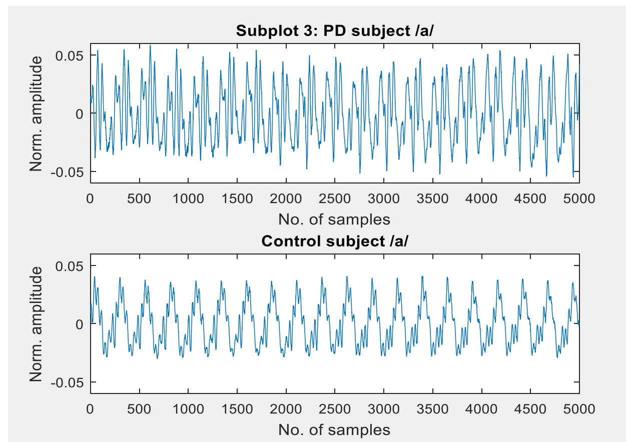

Fig.2. Sustained /a/ for PD and control subject

TABLE 1. FEATURES EXTRACTED FROM THE SUSTAINED PHONEMES

Features
Fundamental frequency (F_0)
Standard deviation of F_0 (SD F_0)
Formant frequencies (F_1, F_2, F_3)
Jitter and shimmer features [15]
Harmonics to noise ratio (HNR) and Noise to Harmonics (NHR)
Soft phonation index (SPI)
Glottal features and glottal based spectral features [16, 17]
Non-linear features [18, 19]
Features based on harmonics (H_1, H_2, H_3) and formant amplitudes (A_1, A_2, A_3)
Mel frequency cepstral coefficients (MFCC) [20]

III. RESULTS AND DISCUSSION

The demographics and clinical information of the subjects are shown in Table 2. In this study, UPDRS-III score was also evaluated for control subjects considering their age-related reduction in motor performance.

TABLE 2. PATIENT DEMOGRAPHICS AND CLINICAL INFORMATION

	PWP (Mean SD)	**Control (Mean SD)**
Age	71.28 (±6.99)	66.91(±6.22)
UPDRS-III (0-132)	25.67(±9.36)	2.64(±3.65)
Duration of PD	4.94(±3.14)	-

To evaluate the efficiency of voice features extracted from sustained consonant /m/ and sustained /a/ in distinguishing PWP and controls, classification was performed independently with reduced features of /a/ and /m/. The results of the SVM classifier is detailed in table 3 below.

TABLE 3. SVM CLASSIFICATION RESULTS FOR SUSTAINED PHONATIONS

Phonation	**CA**	**Sens**	**Spec**	**AUC**	**MCC**
/a/	0.70	0.67	0.73	0.69	0.39
/m/	0.93	0.94	0.91	0.92	0.85

It was observed from the results that the phonetic features extracted from the sustained consonant /m/ differentiated PWP and healthy controls yielding an accuracy rate of 93% in the classification model. This classification accuracy is significantly high compared to that of /a/ which was 70%. For the phonation /m/, MCC was 0.85 and for phonation /a/ it was 0.39. The sensitivity and specificity for the classification model based on /m/ was high compared to that of model based on /a/. The classification results suggested that the classification based on features from phonation /m/ were able to better predict PD. The receiver operating characteristic (ROC) curves for the model based on /a/ and /m/ are shown in Fig.3 and Fig.4.

In our study, the association between UPDRS-III score and the speech features were tested with Spearman correlation coefficient. The association were only tested between the reduced features and UPDRS-III score. The reduced features and the UPDRS-III score were normalised prior to the analysis. The result of correlation analysis is shown in table 4. It was observed that the features from /m/ were strongly correlated ($p < 0.001$) to the UPDRS-III score.

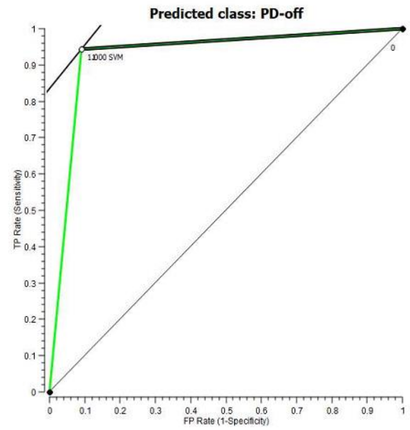

Fig.3. ROC curve from classification model based on reduced features of /m/

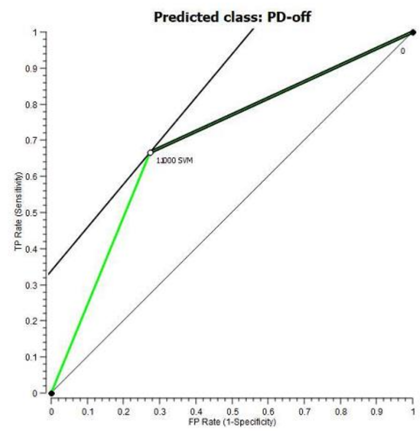

Fig.4. ROC curve from classification model based on reduced features of /a/

TABLE 4. SPEARMAN CORRELATION COEFFICIENT RESULTS

/a/ features	**Spearman correlation coefficient**	**P value**	**/m/ features**	**Spearman correlation coefficient**	**P value**
F_2	0.724	<0.001	F2	0.742	<0.001
F_2-F_1	0.636	<0.001	MFCC C6	0.843	<0.001
Jitter (rap)	0.595	<0.001	Jitter (ppq5)	0.637	<0.001
Jitter (ddp)	0.595	<0.001	MFCC C3	-0.473	0.002
Shimmer (local)	0.338	0.033	Jitter (rap)	0.632	<0.001
Shimmer (local, absolute)	0.366	0.020	Jitter (ddp)	0.632	<0.001
Shimmer (apq5)	0.354	0.025	F_1	0.370	0.019
NHR	0.409	0.009	H_1	0.380	0.016
			MFCC C5	0.680	<0.001

The significance of the current study can be perceived from the results of previous studies carried out with similar study size. Naranjo and colleagues [21] conducted a study to classify PD and control groups by using the features extracted from phoneme /a/ and were able to achieve 75% accuracy in classification. Another study included features from /a/, /e/, /i/, /o/ and /u/ phonemes and obtained classification accuracy ranging from 70-76% and also remarked that the use of features from multiple phonemes did not improve the classification accuracy of the model [22]. The study conducted by Rusz and colleagues utilised three phonemes /a/, /i/, /u/ together with reading tasks concluded that they were able to achieve 80% classification accuracy by using features extracted from the multiple reading tasks [23]. While comparing our results with other studies of similar size as above mentioned, the classification accuracy obtained from features of consonant phoneme /m/ is quite outstanding and promising.

The results from the study indicates the efficiency of using voiced nasal consonant in the diagnosis of PD and it was also evident that the classification accuracy based on /m/ outperformed the model based on /a/ phonation.

The study has certain limitations; one of them being the size of the study population. The second limitation of the study is that due to the limited study size, the effect of gender on the features extracted from the phonemes were not evaluated.

IV. Conclusion

The main objective of the study was to investigate the capability of features extracted from sustained consonant /m/ in distinguishing PD and control groups. Dysphonia features were extracted from /a/ and /m/ phonation and their classification efficiency was studied and compared. The SVM classifier model obtained classification accuracy of 93% with the features extracted from phonation /m/. This result will shift the attention of researchers from using vowel phonations to the consonants as the complexity involved in their production is high. The results can further find application in diagnostics and progression monitoring for PD.

References

[1] G. DeMaagd and A. Philip, "Parkinson's Disease and Its Management: Part 1: Disease Entity, Risk Factors, Pathophysiology, Clinical Presentation, and Diagnosis," *P T,* vol. 40, no. 8, pp. 504-32, Aug 2015.

[2] A. Tsanas, "Accurate telemonitoring of Parkinson’s disease symptom severity using nonlinear speech signal processing and statistical machine learning," DPhil, University of Oxford, 2012.

[3] M. Liotti *et al.*, "Hypophonia in Parkinson's disease: neural correlates of voice treatment revealed by PET," *Neurology,* vol. 60, no. 3, pp. 432-40, Feb 11 2003.

[4] L. Brabenec, J. Mekyska, Z. Galaz, and I. Rektorova, "Speech disorders in Parkinson’s disease: early diagnostics and effects of medication and brain stimulation," *Journal of Neural Transmission,* vol. 124, no. 3, pp. 303-334, 2017.

[5] K. Tjaden, "Speech and Swallowing in Parkinson's Disease," *Top Geriatr Rehabil,* vol. 24, no. 2, pp. 115-126, 2008.

[6] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman, and L. O. Ramig, "Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease," *IEEE Transactions on biomedical engineering,* vol. 59, no. 5, pp. 1264-1271, 2012.

[7] M. Hariharan, K. Polat, and R. Sindhu, "A new hybrid intelligent system for accurate detection of Parkinson's disease," *Computer methods and programs in biomedicine,* vol. 113, no. 3, pp. 904-913, 2014.

[8] M. Peker, B. Sen, and D. Delen, "Computer-aided diagnosis of Parkinson’s disease using complex-valued neural networks and mRMR feature selection algorithm," *Journal of healthcare engineering,* vol. 6, no. 3, pp. 281-302, 2015.

[9] M. Nespor, M. Peña, and J. Mehler, "On the different roles of vowels and consonants in speech processing and language acquisition," *Lingue e linguaggio,* vol. 2, no. 2, pp. 203-230, 2003.

[10] T. Antolík and C. Fougeron, "Consonant distortions in dysarthria due to Parkinson's disease, Amyotrophic Lateral Sclerosis and Cerebellar Ataxia," in *Interspeech 2013,* 2013.

[11] A. Lowit, "Quantification of rhythm problems in disordered speech: a re-evaluation," *Philosophical Transactions of the Royal Society B: Biological Sciences,* vol. 369, no. 1658, p. 20130404, 2014.

[12] K. M. Kurowski, S. E. Blumstein, C. L. Palumbo, R. S. Waldstein, and M. W. Burton, "Nasal consonant production in Broca's and Wernicke's aphasics: speech deficits and neuroanatomical correlates," *Brain Lang,* vol. 100, no. 3, pp. 262-75, Mar 2007.

[13] J. A. Whitfield, A. Reif, and A. M. Goberman, "Voicing contrast of stop consonant production in the speech of individuals with Parkinson disease ON and OFF dopaminergic medication," *Clinical Linguistics & Phonetics,* pp. 1-8, 2017.

[14] D. Cushnie-Sparrow, S. Adams, T. Knowles, T. M. Leszcz, and M. Jog, "Effects of Multi-talker Noise on the Acoustics of Voiceless Stop Consonants in Parkinson's Disease," *Western Papers in Linguistics/Cahiers linguistiques de Western,* vol. 3, no. 1, p. 2, 2016.

[15] D. W. Paul Boersma (22 january). *Praat: doing phonetics by computer*. Available: http://www.praat.org./

[16] Y. D. Heman-Ackah *et al.*, "Quantifying the cepstral peak prominence, a measure of dysphonia," *J Voice,* vol. 28, no. 6, pp. 783-8, Nov 2014.

[17] N. D and N. S. V, "Cepstral Analysis of Voice in Individuals with Parkinson's Disease," in *Proceedings of the International Symposium “Frontiers of Research on Speech and Music (FRSM-2016)*, 2016.

[18] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," *IEEE Trans Biomed Eng,* vol. 56, no. 4, p. 1015, Apr 2009.

[19] I. Hertrich, W. Lutzenberger, S. Spieker, and H. Ackermann, "Fractal dimension of sustained vowel productions in neurological dysphonias: an acoustic and electroglottographic analysis," *The Journal of the Acoustical Society of America,* vol. 102, no. 1, pp. 652-654, 1997.

[20] T. Kapoor and R. Sharma, "Parkinson’s disease diagnosis using Mel-frequency cepstral coefficients and vector quantization," *International Journal of Computer Applications,* vol. 14, no. 3, pp. 43-46, 2011.

[21] L. Naranjo, C. J. Pérez, Y. Campos-Roca, and J. Martín, "Addressing voice recording replications for Parkinson’s disease detection," *Expert Systems with Applications,* vol. 46, pp. 286-292, 2016.

[22] J. R. Orozco-Arroyave, J. D. Arias-Londono, J. F. Vargas-Bonilla, and E. Nöth, "Analysis of speech from people with Parkinson’s disease through nonlinear dynamics," in *International Conference on Nonlinear Speech Processing*, 2013, pp. 112-119: Springer.

[23] J. Rusz *et al.*, "Imprecise vowel articulation as a potential early marker of Parkinson's disease: Effect of speaking task," *The Journal of the Acoustical Society of America,* vol. 134, no. 3, pp. 2171-2181, 2013.