Acoustic analysis of sustained vowels in Parkinson's disease

ACOUSTIC ANALYSIS OF SUSTAINED VOWELS IN PARKINSON'S DISEASE: NEW INSIGHTS INTO THE DIFFERENCES OF UK- AND US-ENGLISH SPEAKING PARTICIPANTS FROM THE PARKINSON'S VOICE INITIATIVE

A. Tsanas[1], S. Arora [2]

[1] Usher Institute, Medical School, University of Edinburgh, Edinburgh, UK
[2] Mathematical Institute, University of Oxford, Oxford, UK
(A. Tsanas): atsanas@ed.ac.uk, tsanasthanasis@gmail.com; (S. Arora): arora@maths.ox.ac.uk

Abstract: **Sustained vowels have often been used to clinically assess vocal performance and infer symptoms in Parkinson's disease, with most studies focusing on cohorts from a single linguistic background. Arguably, sustained vowels are generic and language-independent, however it is not clear how findings might generalize across cohorts of people from different linguistic backgrounds. In this study, we aimed to compare phonations from UK- and US-English speaking people with Parkinson's disease using the largest known speech-Parkinson's database collected using a standard telephone network, the Parkinson's Voice Initiative (PVI). We processed 1988 sustained vowel /a/ phonations from the US-cohort and 525 phonations from the UK-cohort. We stratified data according to gender and computed the fundamental frequency (F0) as a function of age and characterized phonations using 307 acoustic measures that we have used in previous related work. There was generally very good agreement between UK- and US-English speakers in terms of F0 characteristics and traditional acoustic measures such as jitter and shimmer. However, we find pronounced cohort differences with a few of the complex nonlinear acoustic measures. These findings provide useful insights into the acoustic differences between two English speaking cohorts, which should be taken into account when generalizing findings.**

Keywords: **acoustic analysis, Parkinson's disease, speech signal processing, sustained vowels**

I. INTRODUCTION

Parkinson's Disease (PD) is a debilitating progressive neurodegenerative disorder with cornerstone symptoms which include tremor, rigidity and bradykinesia, within the broader remit of motor and non-motor symptoms [1]. PD incidence and prevalence rates have been consistently growing where there was an estimated 6.1 million of People with Parkinson's (PwP) in 2016, and this number is projected to grow further as the average life expectancy increases [2]. Vocal impairment is very common in PD [3] and is met in approximately 70-90% PwP [3].

Studies over the last two decades have demonstrated the enormous potential that speech signals have in neurodegenerative applications including PD. Indicatively, we had previously used sustained vowel /a/ phonations and demonstated: (i) differentiating PwP from age- and gender-matched controls with almost 99% accuracy [4], (ii) accurately replicating the gold standard PD symptom severity score with accuracy greater than the inter-rater variability [5]–[9], and (iii) automatically assisting voice rehabilitation [10]. More recently, we reported on the potential of speech signals towards early PD diagnosis both when using information with *LRRK*2 gene mutations [11] and also with known disease precursors such as rapid eye movement sleep behavior disorder [12]. Similarly, we have developed speech articulation kinematic models to characterize PD dysarthria to provide mechanistic insights into the underlying physiology [13]–[15], and explored PD subgroups [16], [17].

The use of sustained vowels towards the assessment of vocal performance has been well established [18] and in particular towards assessing neurodegenerative disorders [18], [19]. Most studies in the PD research literature focus on cohorts from a single linguistic background, e.g. US-English speakers. Although it could be argued that sustained vowels may be language-independent, there has not been a systematic investigation into acoustic characterization in PwP cohorts from different linguistic backgrounds. This may limit potential comparisons and insights which could be drawn when comparing PwP from different linguistic backgrounds. Motivated by the need to assess speech-PD at large, we initiated the Parkinson's Voice Initiative (PVI), an international study that collected sustained vowel /a/ phonations and basic demographic information from approximately 10,000 people [20]–[22]. This is the largest known speech-PD database and provides a unique opportunity for forming new hypotheses and exploratory analyses.

Please cite this paper as: A. Tsanas, S. Arora: *Acoustic analysis of sustained vowels in Parkinson's disease: new insights into the differences of UK- and US-English speaking participants from the Parkinson's voice initiative*, **12th International Workshop Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA)**, pp. 161-164, Florence, Italy, 14-16 December 2021

In this study, we aimed to compare PwP from UK- and US-English speaking linguistic backgrounds across a range of acoustic characteristics to investigate alignment at a cohort level using age and gender stratification.

II. METHODS

A. Data

The study makes secondary analysis of the PVI data focusing on the UK- and US-English speaking cohorts. We processed 1988 sustained vowel /a/ phonations from the US-cohort and 525 phonations from the UK-cohort. The speech recordings were sampled at 8 kHz and were stored at secure Aculab servers, along with basic demographic information (age, gender). For further details on PVI please see our previous work [20]–[22].

B. Acoustic analysis of sustained vowels

We computed the fundamental frequency (F0) contour using SWIPE [23], which we had previously demonstrated is very competitive in accurate F0 estimation specifically for sustained /a/ vowels [24]. We also used the Voice Analysis Toolbox (MATLAB open source code: https://www.darth-group.com/software), which provides an overview of acoustic characterization of sustained vowels across 307 acoustic measures. These have been specifically developed for PD applications [5], [6], [19], [25] and were later shown to be more broadly applicable to other settings including general voice pathology assessment [26] and forensic phonetics [27]. We compared the UK and US-English speaking cohorts in terms of average F0 and F0 trajectories stratified by age and gender to objectively illustrate overall cohort differences. Also, we compared the cohort distributions across the computed 307 acoustic measures to demonstrate how well these align in the two PwP groups.

III. RESULTS

Fig. 1 presents the average estimated F0 as a function of age, where results are stratified by gender. We observe that the general trend is similar for both cohorts,

Table 1: Indicative acoustic measures of people with Parkinson's, stratified by gender

Acoustic measure	**Brief explanation**	**US cohort (males)**	**US cohort (females)**	**UK cohort (males)**	**UK cohort (females)**
Mean F0	Mean fundamental frequency (F0) computed using SWIPE	139.61±34.03	206.84±33.24	139.17±33.79	216.25±32.98
Jitter	Average successive F0 differences (10 ms windows)	0.49±1.35	0.23±0.64	0.43±1.29	0.21±0.54
Shimmer	Average successive amplitude differences (10 ms windows)	0.10±0.04	0.09±0.04	0.09±0.04	0.10±0.05
NHR	Noise-to-harmonics ratio	0.10±0.24	0.05±0.16	0.06±0.09	0.04±0.14
GNE	Glottal to noise excitation (assessing SNR)	0.88±0.17	1.08±0.21	0.86±0.11	1.09±0.20
$VFER_{mean}$	Vocal fold excitation ratio, average frequency excitation	2.18±2.49	0.95±3.05	2.25±2.34	1.36±3.40
$VFER_{SNR-TKEO}$	Vocal fold excitation ratio, SNR energy excitation	257.40±473.70	313.12±519.29	677.63±835.43	885.88±823.73
PPE	Pitch period entropy (assessing F0 variability)	0.05±0.10	0.02±0.06	0.03±0.08	0.02±0.06
0^{th} MFCC	0th Mel Frequency Cepstral Coefficient	0.92±2.28	1.18±2.24	-0.30±2.11	0.04±2.01
1^{st} MFCC	1st Mel Frequency Cepstral Coefficient	2.10±1.74	1.32±1.67	3.97±1.69	3.40±1.28
12^{th} MFCC	12th Mel Frequency Cepstral Coefficient	0.10±0.40	-0.57±0.47	0.22±0.40	-0.28±0.49

Distributions are summarized in the form mean ± standard deviation. GNE = Glottal to Noise Excitation, MFCC = Mel Frequency Cepstral Coefficient, SNR = Signal to Noise Ratio, VFER = Vocal Fold Excitation Ratio.

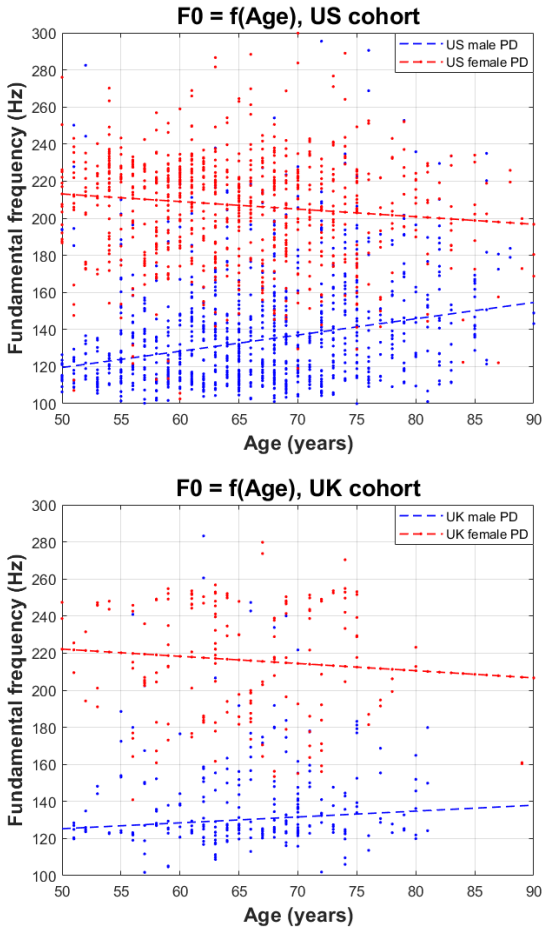

Fig. 1: *Fundamental frequency (F0) as a function of age, stratified by gender for the UK and US PwP cohorts. The best line was computed using robust linear regression fit with iteratively reweighted least squares.*

where the average F0 is increasing with age. However, for both male and female PwP the US cohort exhibit a sharper rate of change.

Table 1 summarizes indicative acoustic measures of the two PwP cohorts to facilitate a side-by-side comparison, stratified by gender. We remark that the classical acoustic measures (e.g. jitter, shimmer, NHR) were very similar. However, there were subtle and pronounced differences in some acoustic measures, in particular the Vocal Fold Excitation Ratio (VFER) measures and Mel Frequency Cepstral Coefficient (MFCC) measures.

IV. Discussion

This study investigated the use of sustained vowel /a/ phonations between speakers from UK- and US-English linguistic backgrounds across a range of acoustic measures. Overall, there was generally very good agreement between the two cohorts in terms of F0 characteristics and most of the acoustic measures investigated. This is a strong indication that clinical decision support tools developed using sustained vowel /a/ phonations in English-speaking PwP cohorts should in principle generalize to other English-speaking PwP. However, there are some subtle pronounced cohort differences with some of the acoustic measures (VFER, MFCCs), which need to be considered when generalizing findings across cohorts with different linguistic backgrounds.

VFER and MFCCs have been particularly successful in related PD clinical decision support tools that we had previously reported on using either UK- or US-English speaking cohorts [7], [12], [19]. The present study's findings could indicate that clinical decision support tools developed across either PwP cohort might need some careful tuning to be generalizable, for example exploring options with transfer learning. In turn, this could also inherently suggest that the PVI cohorts (data collected across seven countries) should be investigated separately to report on individual cohort properties and provide a cross-linguistic comparison of acoustic measure outputs and F0 changes as a function of age.

V. Conclusion

Collectively, these findings support the use of sustained vowels towards vocal assessment in PD as a robust and broadly generalizable signal modality, at least in the English-speaking cohorts. However, care needs to be exercised with some of the acoustic measures (VFER, MFCCs) which appear to differ considerably between cohorts.

References

[1] B. R. Bloem, M. S. Okun, and C. Klein, "Parkinson's disease," *Lancet*, vol. 12, pp. 2284–2303, 2021

[2] E. R. Dorsey *et al.*, "Global, regional, and national burden of Parkinson's disease, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016," *Lancet Neurol.*, vol. 17, pp. 939–953, 2018

[3] 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, no. 3, pp. 131–137, 1998

[4] A. Tsanas *et al.*, "Novel speech signal processing algorithms for high-accuracy classification of Parkinsons disease," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 5, pp. 1264–1271, 2012

[5] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, "Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity.," *J. R. Soc. Interface*, vol. 8, no. 59, pp. 842–

855, 2011
[6] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, "New nonlinear markers and insights into speech signal degradation for effective tracking of Parkinson's disease symptom severity," in *International symposium on nonlinear theory and its applications (NOLTA)*, 2010, pp. 457–460.
[7] A. Tsanas, M. A. Little, and L. O. Ramig, "Remote assessment of Parkinson's disease symptom severity using the simulated cellular mobile telephone network," *IEEE Access*, vol. 9, pp. 11024–11036, 2021
[8] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, "Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 4, pp. 884–893, 2010
[9] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, "Enhanced classical dysphonia measures and sparse regression for telemonitoring of Parkinson's disease progression," *2010 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, pp. 594–597, 2010
[10] A. Tsanas, M. A. Little, C. Fox, and L. O. Ramig, "Objective automatic assessment of rehabilitative speech treatment in Parkinson's disease," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 1, pp. 181–190, 2014
[11] S. Arora *et al.*, "Investigating Voice as a Biomarker for leucine-rich repeat kinase 2-Associated Parkinson's Disease," *J. Parkinsons. Dis.*, vol. 8, no. 4, pp. 503–510, 2018
[12] S. Arora, C. Lo, M. Hu, and A. Tsanas, "Smartphone speech testing for symptom assessment in rapid eye movement sleep behavior disorder and Parkinson's disease," *IEEE Access*, vol. 9, pp. 44813–44824, 2021
[13] A. Gómez *et al.*, "A Neuromotor to Acoustical Jaw-Tongue Projection Model With Application in Parkinson ' s Disease Hypokinetic Dysarthria," *Front. Hum. Neurosci.*, vol. 15, p. 622825, 2021
[14] P. Gómez-Vilda *et al.*, "Phonation biomechanics in quantifying parkinson's disease symptom severity," in *Recent Advances in Nonlinear Speech Processing (Vol. 48 of the series Smart Innovation, Systems and Technologies)*, vol. 48, pp. 93–102, 2016
[15] A. Gómez, A. Tsanas, P. Gómez, D. Palacios-Alonso, V. Rodellar, and A. Álvarez, "Acoustic to kinematic projection in Parkinson's disease dysarthria," *Biomed. Signal Process. Control*, vol. 66, p. e102422, 2021, doi: 10.1016/j.bspc.2021.102422.
[16] A. Tsanas and S. Arora, "Assessing Parkinson's disease speech signal generalization of clustering results across three countries: Findings in the Parkinson's voice initiative study," *BIOSIGNALS 2021 14th Int. Jt. Conf. Biomed. Eng. Syst. Technol. (BIOSTEC 2021)*, pp. 124–134, 2021
[17] A. Tsanas and S. Arora, "Large-scale clustering of people diagnosed with Parkinson's disease using acoustic analysis of sustained vowels: Findings in the Parkinson's voice initiative study," *BIOSIGNALS 2020 13th Int. Jt. Conf. Biomed. Eng. Syst. Technol. (BIOSTEC 2020)*, pp. 369–376, 2020
[18] I. R. Titze, *Principles of voice production*. Iowa City: National Center for Voice and Speech, 2000.
[19] A. Tsanas, "Accurate telemonitoring of Parkinson's disease using nonlinear speech signal processing and statistical machine learning," DPhil thesis, Oxford Centre for Industrial and Applied Mathematics, University of Oxford, 2012.
[20] S. Arora, L. Baghai-Ravary, and A. Tsanas, "Developing a large scale population screening tool for the assessment of Parkinson ' s disease using telephone-quality voice," *J. Acoust. Soc. Am.*, vol. 145, no. 5, pp. 2871–2884, 2019.
[21] A. Tsanas and S. Arora, "Biomedical speech signal insights from a large scale cohort across seven countries: The Parkinson's voice initiative study," in *Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA)*, 2019, pp. 45–48.
[22] A. Tsanas and S. Arora, "Exploring telephone-quality speech signals towards parkinson's disease assessment in a large acoustically non-controlled study," *19th IEEE International Conference on Bioinformatics and Bioengineering*, pp. 953-956, 2019
[23] A. Camacho and J. G. Harris, "A sawtooth waveform inspired pitch estimator for speech and music.," *J. Acoust. Soc. Am.*, vol. 124, no. 3, pp. 1638–52, Sep. 2008
[24] A. Tsanas, M. Zañartu, M. A. Little, C. Fox, L. O. Ramig, and G. D. Clifford, "Robust fundamental frequency estimation in sustained vowels: detailed algorithmic comparisons and information fusion with adaptive Kalman filtering.," *J. Acoust. Soc. Am.*, vol. 135, no. 5, pp. 2885–901, 2014
[25] A. Tsanas, "Acoustic analysis toolkit for biomedical speech signal processing: concepts and algorithms," in *8th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA)*, 2013, pp. 37–40.
[26] A. Tsanas and P. Gómez-Vilda, "Novel robust decision support tool assisting early diagnosis of pathological voices using acoustic analysis of sustained vowels," in *Multidisciplinary Conference of Users of Voice, Speech and Singing (JVHC 13)*, 2013, pp. 3–12.
[27] E. San Segundo, A. Tsanas, and P. Gomez-Vilda, "Euclidean Distances as measures of speaker similarity including identical twin pairs: A forensic investigation using source and filter voice characteristics," *Forensic Sci. Int.*, vol. 270, pp. 25–38, 2017