Comparing lab-based and telephone-based speech recordings towards Parkinson's assessment: insights from acoustic analysis

Please cite this as: A.Tsanas: "Comparing lab-based and telephone-based speech recordings towards Parkinson's assessment: insights from acoustic analysis", *45th IEEE International Conference on Telecommunications and Signal Processing*, pp.55-58, 2022

Comparing lab-based and telephone-based speech recordings towards Parkinson's assessment: insights from acoustic analysis

Athanasios Tsanas
Usher Institute, University of Edinburgh, Edinburgh, UK
atsanas@ed.ac.uk; ORCID: 0000-0002-0994-8100

***Abstract*—The use of high-quality lab-based speech recordings has led to key breakthroughs in a range of Parkinson's Disease (PD) assessment applications. We recently reported on the Parkinson's Voice Initiative (PVI) study collecting telephone-based speech recordings under non-controlled acoustic conditions towards large-scale PD assessment. In this study, we aim to compare the underlying acoustic properties of the sustained vowel /a/ recordings across two large PD datasets focusing only on US speakers to avoid any linguistic confounders. We acoustically characterized 2097 sustained vowel /a/ recordings from 1138 PD participants and compare findings against a large public high-quality speech-PD database of 5875 recordings across 16 dysphonia measures using the symmetric Kullback-Leibler divergence. We explored gender stratification and two-dimensional projections using t-distributed Stochastic Neighbor Embedding (t-SNE) to facilitate visual examination and understand database differences. We find that there are considerable differences in the distributions of the dysphonia measures both univariately and when considered in lower dimensional t-SNE projections even for the linear dysphonia measures. Collectively, these findings provide new insights into understanding the inherent challenges when aiming to generalize findings from lab-based settings to real-world practical applications towards speech-PD clinical decision support tools and may motivate the development of new speech signal processing algorithms.**

***Keywords*—Acoustic characterization, clinical decision support tool, Parkinson's Disease (PD), speech signal processing, tele-assessment**

I. Introduction

Parkinson's Disease (PD) is a progressive neurodegenerative disorder with alarmingly increasing prevalence rates: it is estimated there were approximately 2.5 million people diagnosed with PD (PwP) in 1990, steeply increasing to 6.1 million PwP in 2016 [1]. Strikingly, a large global burden of disease study in 2021 highlighted PD as one of the top five leading causes of death from neurological disorders in the US [1]. Cardinal PD symptoms include tremor, rigidity, bradykinesia, and postural instability, amongst other motor and non-motor symptoms [2]. Crucially for the purposes of this study, PwP also experience considerable speech performance degradation as a key PD symptom [2]–[4].

Given that speech signals are easy to (self-) collect, this has attracted considerable research interest in the PD literature [3], [4]. Typically, sustained vowel /a/ phonations are used in speech clinical assessments because they overcome linguistic effects [3]. Indicative speech-PD applications include: (a) differentiating PwP from age- and gender-matched controls with almost 99% accuracy [5]; (b) accurately replicating the standard clinical scale denoting PD symptom severity [6]–[10]; (c) assessing voice rehabilitation in PD home monitoring systems [11]; (d) providing early PD precursors which may lead to early accurate diagnosis [12], [13]; (e) clustering PwP towards identifying PD subtypes and developing more targeted approaches for individuals [14]; and (f) speech articulation kinematic models to characterize PD dysarthria, thus providing tentative insights into the underlying physiology [15].

Most speech-PD studies report findings by processing recordings collected using high-quality equipment and under highly controlled acoustic conditions. Whilst this is the right first step to demonstrate feasibility under favorable conditions and minimize heterogeneity, current findings may be challenging to scale up in practice if there is a strict requirement of lab-based conditions. To enable large scale analysis into PD we set up a large international multi-site trial, the Parkinson's Voice Initiative (PVI), collecting more than 19,000 sustained vowel /a/ recordings over the standard telephone network, with PwP across 7 countries [16]–[18]. Although the data collected in PVI is not of the same high quality as data collected under carefully controlled acoustic conditions in the lab, the large number of samples facilitates new explorations. However, the reduced data quality poses new challenges, and current algorithms have led to considerable performance degradation in the binary differentiation of PwP and controls [16], [18] compared to the very promising results using the exact data processing methodology we have previously reported when processing lab-based data [5].

Therefore, the motivation for this study was to understand the underlying reasons for those differences in resulting accuracies which might inform further developments towards extracting more nuanced information from the PVI speech recordings. Within this context, the aim of this study was to explore similarities and differences in the acoustic characteristics of the sustained vowels between PVI and a benchmark high-quality database we have previously used [4].

II. Data

The study uses two databases: the Intel At-Home Testing Device (AHTD) [4], [6] and the PVI [16]–[18]. The former used a high-quality device where 42 PwP in the US collected data at home weekly for six months, giving rise to 5875 sustained vowel /a/ phonations. The 42 PwP (28 males, age 64.4±9.24 years) had a PD diagnosis within the previous five years at trial onset and remained un-medicated for the duration of the study. The extracted acoustic characteristics (see III.B) of AHTD is available from the UCI ML repository: http://archive.ics.uci.edu/ml/datasets/Parkinsons+Telemonitoring.

The PVI study invited participants across seven major geographical locations (Argentina, Brazil, Canada, Mexico, Spain, USA, and the UK), to self-enroll and donate their voices towards facilitating PD research. Apart from basic demographics and sustained vowel /a/ phonations, no further information regarding PD onset and symptoms was collected in order to minimize participant burden. In this study we only used the phonations from US participants to minimize confounding factors when comparing to the AHTD dataset: in total, 2097 sustained vowel phonations from 1138 PwP (605 males; age 63.7±10.8 years) were processed for PVI.

For further details on the two databases please refer to the cited studies above where the data was first used.

III. Methods

A. Data preprocessing

In AHTD the authors had developed an automated speech signal processing tool to screen out erroneous recordings (e.g. coughing, non-voiced recordings) [4], and therefore the public version of the dataset used here had already been cleaned. Similarly, in PVI we extended that tool to identify suspicious recordings which do not conform to the expected time-series pattern which were subsequently aurally inspected and discarded as required: for details please see [16].

B. Acoustic characterisation of sustained vowels

To directly compare against the publicly available AHTD dataset we extracted the same 16 dysphonia measures in the PVI dataset (see Table I). We used the implementation that was previously used to generate the AHTD acoustic dataset [4], [6], [19]. The MATLAB source code for the computation of a large range of dysphonia measures is freely available from: https://www.darth-group.com/software.

C. Statistical exploration and dataset comparisons

As a first step we standardized all dysphonia measures, separately in each dataset, to ensure we operate on similar scales. Subsequently, we estimated densities using kernel density estimation with Gaussian kernels for each of the dysphonia measures for the two datasets and present distributions using violin plots. We quantified differences in the distributions of the dysphonia measures between the two datasets using the symmetric Kullback-Leibler Divergence (KLD) [20], which was computed using trapezoidal numerical integration. Moreover, we employed two-dimensional projections using t-distributed Stochastic Neighbor Embedding (t-SNE) [21] to facilitate visual examination of the dysphonia measures *jointly* in the projected two-dimensional space and intuitively understand overall differences. For the optimization of the t-SNE hyper-parameters we followed the methodology we previously outlined in similar applications for the lower dimensional representation [22]. We also explored data stratification by gender, following recommendations in the PD literature [4], [7] and wider speech signal assessment literature [3], [23].

Table I: Summary of dysphonia measures

Measure	**Description**
Jitter(%)	Jitter as a percentage, quantifying changes in fundamental frequency (F0)
Jitter(Abs)	Absolute jitter in microseconds
Jitter: RAP	Relative Amplitude Perturbation
Jitter: PPQ5	Five-point Period Perturbation Quotient
Jitter: DDP	Average absolute differences between cycles, divided by the average period
Shimmer	Quantifying changes in amplitude
Shimmer (dB)	Local shimmer in decibels
Shimmer: APQ3	Three-point Amplitude Perturbation Quotient
Shimmer: APQ5	Five point Amplitude Perturbation Quotient
Shimmer: APQ11	11-point Amplitude Perturbation Quotient
Shimmer: DDA	Average absolute difference between consecutive differences of amplitudes
NHR	Noise-to-Harmonics Ratio
HNR	Harmonics-to-Noise Ratio
RPDE	Recurrence Period Density Entropy
DFA	Detrended Fluctuation Analysis
PPE	Pitch Period Entropy

IV. Results

Figure 1 presents the density distributions summarized using violin plots across the 16 dysphonia measures for both datasets. By visual inspection we observe there are some clear differences between AHTD and PVI in terms of spread. This was verified with the computation of the symmetric KLD which was 4.4 for Jitter: PPQ5, 3.4 for DFA and 1.85 for PPE. The symmetric KLD was considerably smaller (<0.4) for all shimmer variants, NHR and HNR, and RPDE.

Next, we explored the two dimensional representation of the two datasets when considered jointly to visually inspect homogeneity and whether the AHTD and PVI datasets have overall similar properties (see Fig. 2). Whereas the AHTD data are homogeneous, the PVI dataset is more fragmented and a large number of samples appear to be beyond the range of the main bulk of the AHTD data, which tentatively suggests that it spans on a feature space area not represented in AHTD.

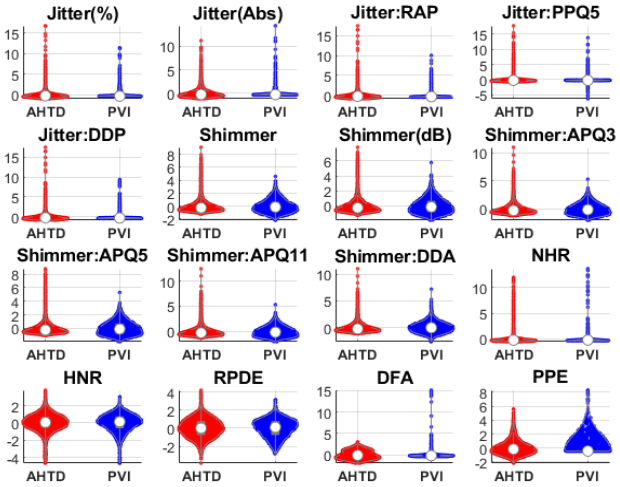

Fig. 1 Density distributions of the computed 16 dysphonia measures across the two databases presented using violin plots for easier visual comparison.

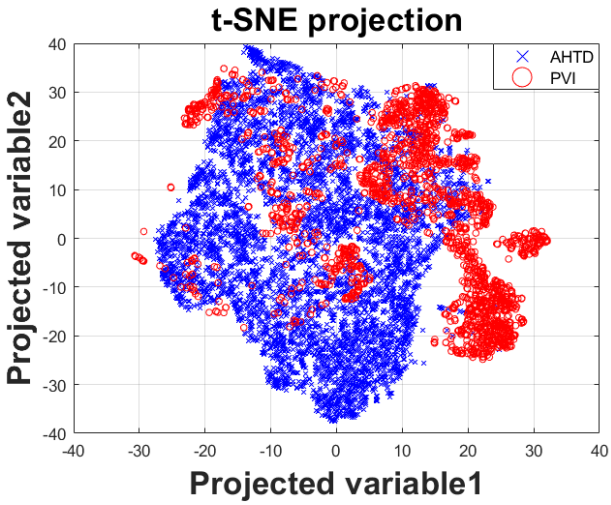

Fig. 2 Two dimensional representation of the AHTD and PVI datasets using t-SNE to explore data homogeneity and differences across datasets. The axes denote the t-SNE projected variables (embedded space) for the two-dimensional data representation.

V. Discussion

We investigated indicative acoustic characteristics of two large databases comparing the AHTD data, where high-quality speech recordings had been collected, and the PVI data, where telephone-based speech recordings had been collected. Both datasets have been collected to enable large scale PD assessment, and we found that there are substantial differences both in stand-along dysphonia measures (see Fig. 1) and in the overall joint representation when projecting the dysphonia measures in a two dimensional feature space (see Fig. 2). These findings suggest that there are inherent differences in the acoustic characteristics between the two databases and tentatively highlight challenges associated with telephone-based quality recordings. Moreover, these exploratory results serve to justify the differences observed in PD assessment with high-quality speech data and telephone-quality speech data. The two dimensional representation (Fig. 2) is revealing of the underlying differences between AHTD and PVI: a large proportion of the PVI samples appear to populate feature space not represented in AHTD. In turn, this finding along with the univariate distributions in Fig. 1 highlights important differences in the acoustic properties of AHTD and PVI which likely reflects the underlying differences of high-quality speech recordings and telephone-based speech recordings. Therefore, this implicitly suggests that dysphonia measures that operate well on high-quality data may not be optimally suited for PVI.

We note that the AHTD and PVI datasets are well age-matched which mitigates the challenge that acoustic characteristic differences might be due to vocal changes as a result of presbyphonia [3]. The differences between the dysphonia measures' distributions were quantified with the symmetric KLD: it is noteworthy that dysphonia measures which have been very successful in different PD applications such as DFA and PPE [4], [7] were markedly different in the two datasets. As part of our exploration we also worked on gender stratified datasets and observed that results did not substantially differ from those presented herein; due to space constraints these are not shown.

The choice of the 16 dysphonia measures used in this study was mandated by the dysphonia measures that are available in the public AHTD dataset to enable direct comparisons. A further practical consideration for the purposes of this study is that the dysphonia measures explored here mainly rely on the linear properties of the speech signal and do not inherently require considerable signal bandwidth (or in other words the higher end of the signal spectrum) [3], [4]. This is a particularly important consideration for the choice of dysphonia measures when processing telephone-based speech, where the sampling frequency is 8 kHz: more advanced nonlinear dysphonia measures use spectral information beyond 4 kHz [4], [7] and therefore are fundamentally not well-suited for this type of lower quality speech recordings [16]–[18].

There are certain limitations to the exploratory nature of this work which we acknowledge. Participants in the PVI study were self-selected and reported whether they had a clinical PD diagnosis: there is no way we can verify this was accurate. Moreover, the sampling frequency of the PVI data was 8 kHz, a major limitation for biomedical speech signal analysis where the recommended sampling frequency is 20 kHz [3]. This was imposed by the use of the standard telephone network and was a pragmatic trade-off for the collection of such a large database. Finally, participants used their standard phone devices (landline or mobile phones), which may introduce different signal distortions. On the other hand, for the AHTD study, participants remained off PD-medication and provided phonations on a weekly basis: using data from only 42 participants might not be sufficiently representative as in PVI.

Future work could integrate additional modalities including wearable sensor data [24] and other modalities such as PwP self-reports, e.g. the mPower study [25]. This would enable a large scale multimodal exploration, and we can use advanced machine learning tools to determine the most parsimonious subset of features or modalities towards specific PD applications building on our previous framework [26], [27].

VI. Conclusion

The present study's findings provide further insights to explain differences between the reported high accuracies in the literature using high-quality speech recordings [5] and accuracies reported using the data collected over the standard phone network in PVI [16], [18]. We envisage these considerations may be useful if the field is to move beyond standard highly controlled research studies which are rarely translated into clinical practice [28], and therefore researchers should explore innovative solutions, which come with new challenges, to facilitate uptake and generalize findings at scale.

References

[1] V. L. Feigin *et al.*, "Burden of Neurological Disorders across the US from 1990-2017: A Global Burden of Disease Study," *JAMA Neurol.*, vol. 78, no. 2, pp. 165–176, 2021

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

[3] I. R. Titze, *Principles of voice production*. Iowa City: National Center for Voice and Speech, 2000

[4] A. Tsanas, "Accurate telemonitoring of Parkinson's disease using nonlinear speech signal processing and statistical machine learning," University of Oxford, 2012

[5] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman, and L. O. Ramig, "Novel speech signal processing algorithms for high-accuracy classification of Parkinsons disease," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 5, pp. 1264–1271, 2012

[6] 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, pp. 884–893, 2010

[7] 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

[8] 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, no. September, pp. 457–460

[9] 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, p. 11024–11036, 2021

[10] 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.*, pp. 594–597, 2010

[11] 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, pp. 181–190, 2014

[12] 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

[13] 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

[14] A. Tsanas and S. Arora, "Data-driven subtyping of Parkinson's using acoustic analysis of sustained vowels and cluster analysis: findings in the Parkinson's voice initiative study," *SN Computer Science, Vol. 3:232*, 2022

[15] P. Gómez-Vilda *et al.*, "Phonation biomechanics in quantifying parkinson's disease symptom severity," in *Recent Advances in Nonlinear Speech Processing*, vol. 48, 2016, pp. 93–102

[16] 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

[17] 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

[18] S. Arora and A. Tsanas, "Assessing Parkinson's disease at scale using telephone-recorded speech: insights from the Parkinson's Voice Initiative," *Diagnostics*, vol. 11, no. 10, p. e1892, 2021

[19] 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

[20] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley, 2005

[21] L. van der Maaten and G. Hinton, "Visualizing Data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, 2008

[22] A. P. Vogel, A. Tsanas, and M. L. Scattoni, "Quantifying ultrasonic mouse vocalizations using acoustic analysis in a supervised statistical machine learning framework," *Sci. Rep.*, vol. 9, no. 1, p. e8100, 2019

[23] 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

[24] A. Tsanas, E. Woodward, and A. Ehlers, "Objective characterization of activity, sleep, and circadian rhythm patterns using a wrist-worn actigraphy sensor: insights into post-traumatic stress disorder," *JMIR mHealth uHealth*, vol. 8, no. 4, p. e14306, 2020

[25] B. M. Bot *et al.*, "The mPower study, Parkinson disease mobile data collected using ResearchKit," *Sci. Data*, vol. 3, p. 160011, 2016

[26] A. Tsanas, "Relevance, redundancy and complementarity trade-off (RRCT): a principled, generic, robust feature selection tool," *Patterns*, vol. 3, pp. 100471, 2022

[27] E. Naydenova, *et al.*, "The power of data mining in diagnosis of childhood pneumonia," *J. R. Soc. Interface*, vol. 13, p. 20160266, 2016

[28] A. K. Triantafyllidis and A. Tsanas, "Applications of Machine Learning in Real-life Digital Health Interventions: Review of the Literature," *J. Med. Internet Res.*, vol. 21, no. 4, p. e12286, 2019