The classification of neurodegenerative disease from acoustic speech data

Benjamin Schultz (✉ ben.schultz@unimelb.edu.au)

The University of Melbourne <https://orcid.org/0000-0002-8306-0206>

Zaher Joukhadar

The University of Melbourne <https://orcid.org/0000-0002-8130-270X>

Maria del Mar Quiroga

The University of Melbourne <https://orcid.org/0000-0002-8943-2808>

Usha Nattala

The University of Melbourne

Gustavo Noffs

The University of Melbourne

Sandra Rojas

The University of Melbourne

Hannah Reece

The University of Melbourne

Anneke van der Walt

Monash University

Adam Vogel

The University of Melbourne <https://orcid.org/0000-0002-3505-2631>

Article

Keywords: Neurodegenerative disease, classification, speech acoustics

Posted Date: November 30th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1029846/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Neurodegenerative diseases often affect speech. Speech acoustics can be used as objective clinical markers of pathology. Previous investigations of pathological speech have primarily compared controls with one specific condition and excluded comorbidities. We broaden the utility of speech markers by examining how multiple acoustic features can delineate diseases. We used supervised machine learning with gradient boosting (CatBoost) to differentiate healthy speech and speech from people with multiple sclerosis or Friedreich ataxia. Participants performed a diadochokinetic task where they repeated alternating syllables. We extracted 74 spectral and temporal prosodic features from the speech recordings, which were subjected to machine learning. Results showed that Friedreich ataxia, multiple sclerosis and healthy controls were all identified with high accuracy (over 82%). Twenty-one acoustic features were strong markers of neurodegenerative diseases, falling under the categories of spectral qualia, spectral power, and speech rate. We demonstrated that speech markers can delineate neurodegenerative diseases and distinguish healthy speech from pathological speech with high accuracy. Findings emphasize the importance of examining speech outcomes when assessing indicators of neurodegenerative disease. We propose large-scale initiatives to broaden the scope for differentiating other neurological diseases and affective disorders.

Introduction

Neurodegenerative diseases can impair speech due to the breakdown of motor control. Therefore, acoustic features of speech can be used as objective clinical markers for neurodegenerative disease. Previous studies examining acoustic changes in neurodegenerative disease have primarily focused on differences between healthy controls and various patient populations, such as, multiple sclerosis (MS) (1–4), Huntington's disease (HD) (5–7), Parkinson's disease (PD) (1,3,8,9), and Friedreich ataxia (FA) (8,10,11). These studies report that various acoustic features indeed change as the disease progresses, and patients tend to exhibit slower and more variable speech rates, lower and more variable pitch, and reduced spectral clarity compared to patients. Although machine learning has been used to differentiate healthy controls from single, well-defined patient populations (e.g., Parkinson's disease, spasmodic dysphonia), real-world machine learning implementations will encounter multiple different diseases that may have overlapping acoustic outcomes. The present study aims to broaden the utility of these speech markers by determining how different acoustic profiles of speech may accurately identify specific neurodegenerative diseases. In other words, we advance beyond discriminating between healthy and pathological speech by examining differences between different neurodegenerative diseases simultaneously across multiple acoustic dimensions.

Clinicians use a combination of tools to diagnose neurodegenerative disease including genetic sequencing, neurological scans (e.g., magnetic resonance imaging), and motor tests. Although these methods have improved considerably, misdiagnosis is still possible, particularly during earlier disease stages where motor problems, cognitive impairment, and/or mood disturbances could be misattributed to other neurological conditions (e.g., (12,13)). Acoustic clinical markers have several advantages over

traditional tools. First, speech can be recorded remotely in a home environment without the need to visit a specialist or hospital. Given that some populations with neurodegenerative disease are considered at-risk, remote identification reduces the risk of contracting potentially life-threatening pathogens. Remote testing is also more accessible for populations with limited mobility and those living in rural areas. Second, acoustic markers are obtained using non-invasive techniques. Invasive procedures (e.g., blood tests, surgery) can cause discomfort and have a risk of infection. These procedures can also impose large financial burdens, particularly when multiple tests are required due to misdiagnosis. Acoustic markers have the potential to alleviate these burdens by providing accessible, low-cost, and low-risk tests that can guide clinician decisions in the early stages of diagnosis.

Acoustic Markers Of Pathological Speech

Acoustic features of speech can be used to construct profiles of different patient populations. The most common acoustic features used to identify pathologies include speech rate, the number of syllables per second, pauses, the duration between utterances or syllables, and frequency information related to the pitch of the voice (fundamental frequency; f_0) and its formants (cf. 12). These features reflect underlying dysfunctions of clinical populations that have decreased motor control over articulators during speech. As such, neurodegenerative disease can result in a slower and more variable speech rate, longer pauses, and decreased pitch quality compared to controls. Other spectral features have also been shown to be useful markers of neurodegenerative disease (e.g., jitter and shimmer) as they capture instabilities in pitch and amplitude that arise as a result of tremor or decreased muscle strength within the articulatory system (7).

There are multiple speech elicitation tasks for speech pathology assessment. The diadochokinetic (DDK) task is a common method of speech elicitation that requires the speaker to produce as many repetitions of /PATAKA/ as possible within one breath or for a specified duration (15). The DDK task is a controlled method of speech elicitation that allows high consistency between different speakers in terms of the speech content while remaining sensitive to oromotor performance (e.g., speech rate) (16). Other speech elicitation tasks include reading a paragraph aloud or semi-structured interviews (17,18). Although these tasks increase ecological validity, they may also increase cognitive load, which may induce speech changes based on third variables like education level, language or reading impairments, or other confounds based on cognitive ability. Moreover, the linguistic content may encourage changes in prosodic features based on emphasis, stress patterns, and emotion which may differ based on personality, accent, or emotional state. To avoid these concerns, the present study examined speech from a DDK task that was performed by healthy controls (HC) and two patient populations (FA, MS) using uniform practices (see Methods). We calculated acoustic features that have been examined in previous studies comparing HCs and various patient populations (1–9).

Previous implementations of machine learning on speech have compared healthy controls with only a single patient group (cf. (14)). Although these approaches are useful as initial triage for identifying

pathological voice disturbances that should be investigated (19), they do not provide nuanced classification of the underlying pathology or disease phenotypes potentially due to small sample sizes and, consequently, low accuracy (20). This is especially the case for models with hidden layers that reflect latent variables that are not defined and, therefore, do not aid in developing specific acoustic profiles that characterize a disease (21). We used a transparent machine learning approach using gradient boosting that quantifies the contribution of each acoustic feature in distinguishing between healthy and pathological voices, and between multiple diseases.

Results

Two-sample *t*-tests revealed that overall model performance as assessed by Matthews's correlation coefficient ($M = 0.82$, $SD = 0.04$) was significantly above chance (33% accuracy), $t(99) = 114.56$, $p < 0.001$, Cohen's $d = 11.46$. Classification accuracy between groups was assessed using $F1$ scores that equally weight model specificity and sensitivity (see supplementary materials for full statistical analysis); these were also significantly better than chance for all groups ($ps < 0.001$) with large effect sizes for HC (Cohen's $d = 32.29$), MS (Cohen's $d = 11.82$), and FA (Cohen's $d = 10.70$). There was a significant main effect of group, $F(2, 198) = 238.20$, $p < 0.001$, $\eta^2_G = 0.49$. As shown in Figure 1, classification accuracy was higher for HC compared to FA ($p < 0.001$) and MS ($p < 0.001$), and higher for FA compared to MS ($p < 0.001$). Receiver operating curves (ROC) for the model with average performance based on Matthew's correlation coefficients are shown in Figure 2. The ROC area under the curve (AUC) values were 0.97 for HC, 0.98 for FA, and 0.96 for MS. These values indicate outstanding discrimination by the model (22).

Model optimization

To measure the contribution of each acoustic feature for categorising each group, the Shapley additive explanation (SHAP) values were examined. These show the probability of each outcome based on the information provided by each feature (23, 24). To achieve a more parsimonious model, we performed the same machine learning procedures twice including features that produced SHAP values above criteria of 2% ($n=87$) and 5% ($n=21$) for at least one group (see supplementary materials for rankings). Overall model accuracy (Matthew's Correlation Coefficient) significantly increased relative to the full model ($M = 82.3\%$, $SEM = 0.4\%$) for the 2% cut-off ($p = 0.001$; $M = 83.7\%$, $SEM = 0.4\%$) and 5% cut-off ($p < 0.001$; $M = 83.6\%$, $SEM = 0.4\%$). Pairwise comparisons of accuracy between models for each group revealed significant increases in $F1$ accuracy between the full model and the 2% cut-off for all groups ($ps < .002$), and between the full model and 5% cut-off for the HC and MS group ($ps < 0.03$) but not the FA group ($p = 0.11$) (see Figure 1). These results suggest that high discrimination accuracy can be achieved with a reduced subset of 21 acoustic features. It should be noted, however, that larger subsets of variables may be required to achieve high discrimination accuracy if a broader scope of clinical groups are included.

Optimal acoustic clinical markers

Here we describe the top 21 features overall, and top 10 for each group and overall (see supplementary materials for all features). As shown in Table 1, the dominant acoustic features for accurate classification were spectral decrease, peak f_0 energy, peak energy in the low, high, and broadband frequency ranges, low-frequency summed energy, utterance duration based on summed broadband energy (including low-, mid- and high-frequency sub-bands), spectral spread, and acoustic intensity. Figures 3a and 3b show that healthy controls are characterized by a less steep and less variable spectral decrease, a smaller spectral spread and range of energy produced in low frequencies, greater energy in low and f_0 frequency bands, and shorter utterance durations. The FA group was characterized by low intensity and energy in low, high, and broadband frequency bands, a higher and more variable spectral spread, and longer utterance durations. The MS group was characterized by a steeper and more variable spectral decrease, as well as utterance durations and spectral spread values that fell between the control and FA groups (see link in Figure 3a note for figures of all acoustic features). Other acoustic features that were useful in delineating groups include metrics of speech timing (pause duration, speech rate, and stress rate (25)), spectral features (crest, slope, centroid, flatness, and entropy (26)), formants 1-5, and the alpha ratio (27).

Table 1: *Top 10 acoustic features for categorizing groups based on SHAP values.*

Rank	Overall	Healthy Control (HC)	Friedreich Ataxia (FA)	Multiple sclerosis (MS)
1	Min. Spectral decrease	Min. Spectral decrease	Min. Spectral decrease	Min. Spectral decrease
2	Mean Spectral decrease	Mean Spectral decrease	Min. f_0 peak energy	Mean Spectral decrease
3	Range Spectral decrease	Range Spectral decrease	Min. High-frequency peak energy	Range Spectral decrease
4	Range Low-frequency sum. energy	Range Low-frequency sum. energy	Min. Utterance duration (low-freq. sum. energy)	Range Low-frequency sum. energy
5	Min. High-frequency peak energy	Max. Spectral spread	Max. Spectral decrease	Min. High-frequency peak energy
6	SD Spectral decrease	SD Spectral decrease	Mean Utterance duration (broadband sum. energy)	SD Spectral decrease
7	Max. Spectral spread	Min. Low-frequency peak energy	Mean Utterance duration (mid-freq. sum. energy)	Min. Acoustic intensity (dB SPL)
8	Min. Low-frequency peak energy	Min. Broadband peak energy	Min. Low-frequency peak energy	Max. Spectral spread
9	Min. Acoustic intensity (dB SPL)	Range Spectral spread	Mean High-frequency sum. energy	Min. f_0 peak energy
10	Min. f_0 peak energy	CoV Spectral decrease	Mean Utterance duration (high-freq. sum. energy)	Min. Utterance duration (low-freq. sum. energy)

Discussion

Our machine learning approach was able to distinguish between healthy controls and two cohorts with different neurodegenerative diseases with high accuracy using acoustic properties of speech alone. These results indicate that multiclass supervised machine learning has the potential to discriminate

between diseases, a step beyond the mere healthy-pathological dichotomy. Therefore, through the accumulation of big data that merges speech data from various patient populations, we may be able to use machine learning to assist in the detection of specific diseases using acoustic markers.

There are numerous advantages for using acoustic markers to detect neurodegenerative disease including the decreased risk and burden of travelling to a hospital to undergo a range of tests, some of which are invasive. Speech, on the other hand, can be recorded within a familiar and comfortable setting, using common household devices (e.g., smartphones). Smartphones have demonstrated relative robustness for obtaining acoustic clinical markers and, therefore, increase accessibility to these automated detection methods (28). Although the present study recorded speech within laboratory settings, it is also possible to record speech data remotely (29). Practitioners could use this information to refine which tests should be performed to confirm a diagnosis. This would be particularly useful for people living in rural communities with increased travel burdens or during situations where the risk of infection is heightened (e.g., pandemics). Speech markers can be used as a remote tool to initially detect signs of neurodegenerative disease, expand our understanding of the clinical characteristics of these diseases to improve our ability to develop targeted interventions, and to monitor disease progression.

We identified several acoustic features that strongly contributed to distinguishing between groups. Spectral decrease, the average of all slopes between the peak amplitude at the fundamental frequency and the peak amplitude of the formants (i.e., harmonics), was the most useful variable in distinguishing our three groups. This finding is in line with previous results that suggest vocal fold dysfunctions are associated with greater energy in the lower frequency range relative to higher frequencies (e.g., the soft phonation index (30)). Other spectral features associated with the distribution of vocal energy also contributed to classification accuracy, including summed and peak energy within low-frequency bands (1-75 Hz), peak energy within f_0 (75-500Hz), high (4000-8000 Hz), and broadband (1-8000Hz) frequency ranges, and the spectral spread of peak frequencies. Therefore, the distribution of acoustic energy across the spectrum that reflects voice qualia culminates as a strong set of acoustic clinical markers for distinguishing neurodegenerative diseases.

Speech timing measures were also strong contributors to classification accuracy, specifically, the duration of syllables based on summed energy in low and broadband frequency ranges, and the rate of stressed syllable onsets based on peak energy across broadband frequencies (25). These results corroborate previous findings that demonstrate slowed speech rate and decreased phonation time for a range of neurodegenerative diseases including Parkinson's disease (31), Huntington's disease (6, 32), multiple sclerosis (1, 3), and others (32, 33). Speech rate and phonation time reflect both pneumo-articulatory capacity and oromotor function, and could serve as clinical markers for neurodegenerative diseases and their progression. Therefore, speech timing measures are useful markers for distinguishing between diseases and may also aid in determining disease severity.

Limitations & Considerations

We used the most common acoustic features (or similar proxies) based on an *a priori* analysis of the neurodegenerative disease literature and timbral features used in music information retrieval. There are, however, other acoustic features that may increase the accuracy and sensitivity of the machine learning algorithm that were not considered here. For example, voice onset time, the time between the burst of a stop consonant and the onset of the vowel, is an acoustic feature that differs significantly between controls and people with Parkinson's disease (34). We opted not to use this measure because our data contain a high degree of coarticulation, and there is little agreement for the best way to extract the burst and vowel onset times and which acoustic features should be considered (see (35)). Similarly, we did not include measures from other voice assessment tasks (e.g., sustained vowel) (36) that can more reliably measure certain features (e.g., jitter and shimmer) but preclude the measurement of speech timing. We chose to constrain the number of variables and tasks to avoid overfitting. Future studies could use feature selection and pruning methods (e.g., (37)) to find the best feature set and remove unreliable variables prior to analysis.

The inclusion of non-speech performance measurements could also increase discrimination accuracy, for instance, cognitive (38) and motor performance (39) measures. The primary aim of this experiment was to examine accuracy using speech features alone because speech data can easily be obtained in the absence of a clinician through websites and smartphone applications (40). Other cognitive and motor tests often require the scoring of a clinician or dedicated tools to measure gait and tremor, although some smartphone tests are available (41). We show that neurodegenerative diseases can be delineated with high accuracy from speech data alone, but future applications could also consider other non-verbal features, for example, irregular gait patterns using smartphone accelerometers or irregular typing patterns. Whether these movement features or others would increase the accuracy of machine learning algorithms for neurodegenerative disease remains unknown.

Conclusion

We provide strong evidence that neurodegenerative diseases can be differentiated through acoustic clinical markers and machine learning. This model can be expanded and improved through the inclusion of additional diseases and phenotypes. Big data initiatives that bring together researchers and speech data from multiple laboratories are necessary to increase the scope of diseases that can be identified by acoustic clinical markers and machine learning. Moreover, a combination of remote testing tools for physical and cognitive assessment could be included in addition to speech to improve identification accuracy. These technologies promise to provide tools that can aid practitioners in reaching a diagnosis and relieve the physical and financial burden of patients.

Declarations

Acknowledgements

This work was undertaken in collaboration with the Melbourne Data Analytics Platform (MDAP) at The University of Melbourne. Data collection for the multiple sclerosis group was supported by NHMRC Project grant (#108546). APV was supported by a NHMRC Fellowship (#1135683).

Contributions

BGS managed the machine learning, performed the acoustic feature extraction and statistical analysis, interpreted the data, and wrote the manuscript. ZJ, UN, and MMQ performed the data wrangling and machine learning. Healthy control data were collected by GN and SR under the supervision of APV and AvdW. Data for people with multiple sclerosis were collected by GN. Data for people with Friedreich ataxia were collected by HR and APV. Speech production protocols were designed by APV. BGS and APV contributed to the conception of the study. All authors were involved in manuscript revisions and have approved the final manuscript.

Competing interests

APV is the CEO of Redenlab, a speech clinical marker company.

Materials & correspondence

Raw speech data are not publicly available due to Institutional Review Board restrictions. Participants did not consent to their data being publicly available. Processed data and code used for data processing and analysis is available on request from the corresponding author (BGS). Descriptions of protocols are available on request from the corresponding author (APV).

References

1. Tjaden K, Martel-Sauvageau V. Consonant acoustics in parkinson's disease and multiple sclerosis: Comparison of clear and loud speaking conditions. *Am J Speech-Language Pathol*. 2017;26(2Special Issue):569–82.
2. Feijó AV, Parente MA, Behlau M, Haussen S, De Veccino MC, De Faria Martignago BC. Acoustic analysis of voice in multiple sclerosis patients. *J Voice*. 2004;18(3):341–7.
3. Tjaden K, Lam J, Wilding G. Vowel acoustics in Parkinson's disease and multiple sclerosis: comparison of clear, loud, and slow speaking conditions. *J Speech Lang Hear Res* [Internet]. 2013/07/09. 2013 Oct;56(5):1485–502. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/23838989>
4. Noffs G, Boonstra FMC, Perera T, Butzkueven H, Kolbe SC, Maldonado F, et al. Speech metrics, general disability, brain imaging and quality of life in multiple sclerosis. *Eur J Neurol*. 2021;28(1):259–68.
5. Rusz J, Klempíř J, Baborová E, Tykalová T, Majerová V, Čmejla R, et al. Objective Acoustic Quantification of Phonatory Dysfunction in Huntington's Disease. *PLoS One*. 2013;8(6).

6. Skodda S, Schlegel U, Hoffmann R, Saft C. Impaired motor speech performance in Huntington's disease. *J Neural Transm*. 2014;121(4):399–407.
7. Kaploun LR, Saxman JH, Wasserman P, Marder K. Acoustic analysis of voice and speech characteristics in presymptomatic gene carriers of Huntington's disease: biomarkers for preclinical sign onset? *J Med Speech - Lang Pathol* [Internet]. 2011 Jan 21;19:49+. Available from: <https://link.gale.com/apps/doc/A260582935/PPPM?u=unimelb&sid=PPPM&xid=d1828def>
8. Lansford KL, Liss JM. Vowel acoustics in dysarthria: speech disorder diagnosis and classification. *J Speech Lang Hear Res* [Internet]. 2014 Feb;57(1):57–67. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/24687467>
9. Magee M, Copland D, Vogel AP. Motor speech and non-motor language endophenotypes of Parkinson's disease. *Expert Rev Neurother*. 2019;19(12):1191–200.
10. Vogel AP, Wardrop MI, Folker JE, Synofzik M, Corben LA, Delatycki MB, et al. Voice in Friedreich Ataxia. *J Voice* [Internet]. 2017;31(2):243.e9-243.e19. Available from: <http://dx.doi.org/10.1016/j.jvoice.2016.04.015>
11. Rosen KM, Folker JE, Murdoch BE, Vogel AP, Cahill LM, Delatycki MB, et al. Spectral measures of the effects of Friedreich's ataxia on speech. *Int J Speech Lang Pathol*. 2011;13(4):329–34.
12. Pascu AM, Ifteni P, Teodorescu A, Burtea V, Correll CU. Delayed identification and diagnosis of Huntington's disease due to psychiatric symptoms. *Int J Ment Health Syst* [Internet]. 2015;9(1):33. Available from: <https://doi.org/10.1186/s13033-015-0026-6>
13. Warner J, Barron L, St Clair D, Brock D. Reliability of clinical diagnosis of Huntington's disease. *J Neurol Neurosurg & Psychiatry* [Internet]. 1994 Oct 1;57(10):1277. Available from: <http://jnnp.bmj.com/content/57/10/1277.1.abstract>
14. Hegde S, Shetty S, Rai S, Dodderi T. A Survey on Machine Learning Approaches for Automatic Detection of Voice Disorders. *J Voice* [Internet]. 2019;33(6):947.e11-947.e33. Available from: <http://www.sciencedirect.com/science/article/pii/S0892199718301437>
15. Prins TD. Motor and auditory abilities in different groups of children with articulatory deviations. *J Speech Hear Res*. 1962;5(2):161–8.
16. Ben-David BM, Icht M. The Effect of Practice and Visual Feedback on Oral-Diadochokinetic Rates for Younger and Older Adults. *Lang Speech* [Internet]. 2017 Jun 14;61(1):113–34. Available from: <https://doi.org/10.1177/0023830917708808>
17. Powell TW. A comparison of English reading passages for elicitation of speech samples from clinical populations. *Clin Linguist Phon*. 2006;20(2–3):91–7.

18. Sorensen T, Toutios A, Goldstein L, Narayanan SS. Characterizing Vocal Tract Dynamics Across Speakers Using Real-Time MRI. In: Interspeech. 2016. p. 465–9.
19. Ghoraani B, Krishnan S. A joint time-frequency and matrix decomposition feature extraction methodology for pathological voice classification. EURASIP J Adv Signal Process. 2009;2009:1–11.
20. Jothilakshmi S. Automatic system to detect the type of voice pathology. Appl Soft Comput. 2014;21:244–9.
21. Myszczyńska MA, Ojames PN, Lacoste AMB, Neil D, Saffari A, Mead R, et al. Applications of machine learning to diagnosis and treatment of neurodegenerative diseases. Nat Rev Neurol. 2020;16(8):440–56.
22. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol. 2010;5(9):1315–6.
23. Nohara Y, Matsumoto K, Soejima H, Nakashima N. Explanation of machine learning models using improved Shapley Additive Explanation. In: Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics. 2019. p. 546.
24. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: Proceedings of the 31st international conference on neural information processing systems. 2017. p. 4768–77.
25. Schultz BG, O'Brien I, Phillips N, McFarland DH, Titone D, Palmer C. Speech rates converge in scripted turn-taking conversations. Appl Psycholinguist. 2016;37(5):1201–20.
26. Côrtes Gama AC, Camargo Z, Rocha Santos MA, Carlos Rusilo L. Discriminant Capacity of Acoustic, Perceptual, and Vocal Self: The Effects of Vocal Demands. J Voice [Internet]. 2015;29:260.e45-260.e50. Available from: <https://ezp.lib.unimelb.edu.au/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edselp&AN=S0892199714001313&site=eds-live&scope=site>
27. Vogel AP, Fletcher J, Maruff P. Acoustic analysis of the effects of sustained wakefulness on speech. J Acoust Soc Am. 2010;128(6):3747–56.
28. Jannetts S, Schaeffler F, Beck J, Cowen S. Assessing voice health using smartphones: bias and random error of acoustic voice parameters captured by different smartphone types. Int J Lang Commun Disord. 2019;54(2):292–305.
29. Tsanas A, Little MA, Ramig LO. Remote assessment of Parkinson's disease symptom severity using the simulated cellular mobile telephone network. IEEE Access. 2021;9:11024–36.
30. Bhuta T, Patrick L, Garnett JD. Perceptual evaluation of voice quality and its correlation with acoustic measurements. J Voice [Internet]. 2004;18:299–304. Available from:

<https://ezp.lib.unimelb.edu.au/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edselp&AN=S0892199703001735&site=eds-live&scope=site>

31. Skodda S, Schlegel U. Speech rate and rhythm in Parkinson's disease. *Mov Disord Off J Mov Disord Soc.* 2008;23(7):985–92.
32. Vogel AP, Shirbin C, Churchyard AJ, Stout JC. Speech acoustic markers of early stage and prodromal Huntington's disease: a marker of disease onset? *Neuropsychologia.* 2012;50(14):3273–8.
33. Noffs G, Boonstra FMC, Perera T, Kolbe SC, Stankovich J, Butzkueven H, et al. Acoustic Speech Analytics Are Predictive of Cerebellar Dysfunction in Multiple Sclerosis. *The Cerebellum.* 2020;1–10.
34. Montaña D, Campos-Roca Y, Pérez CJ. A Diadochokinesis-based expert system considering articulatory features of plosive consonants for early detection of Parkinson's disease. *Comput Methods Programs Biomed.* 2018;154:89–97.
35. Auzou P, Özsancak C, Morris RJ, Mary J, Eustache F, Hannequin D. Voice onset time in aphasia, apraxia of speech and dysarthria: A review. *Clin Linguist Phonetics.* 2000;14(2):131–50.
36. Lee L, Stemple JC, Glaze L, Kelchner LN. Quick screen for voice and supplementary documents for identifying pediatric voice disorders. *Lang Speech Hear Serv Sch.* 2004;
37. Zhang Z. Variable selection with stepwise and best subset approaches. *Ann Transl Med.* 2016;4(7):1–6.
38. Jalakas M, Palmqvist S, Hall S, Svärd D, Lindberg O, Pereira JB, et al. A quick test of cognitive speed can predict development of dementia in Parkinson's disease. *Sci Rep.* 2019;9(1):1–8.
39. Montero-Odasso M, Pieruccini-Faria F, Bartha R, Black SE, Finger E, Freedman M, et al. Motor phenotype in neurodegenerative disorders: gait and balance platform study design protocol for the Ontario Neurodegenerative Research Initiative (ONDRI). *J Alzheimer's Dis.* 2017;59(2):707–21.
40. Arora S, Lo C, Hu M, Tsanas A. Smartphone Speech Testing for Symptom Assessment in Rapid Eye Movement Sleep Behavior Disorder and Parkinson's Disease. *IEEE Access.* 2021;
41. McLaren B, Andrews SC, Glikmann-Johnston Y, Mercieca E-C, Murray NWG, Loy C, et al. Feasibility and initial validation of 'HD-Mobile', a smartphone application for remote self-administration of performance-based cognitive measures in Huntington's disease. *J Neurol.* 2021;268(2):590–601.
42. Neto EC, Pratap A, Perumal TM, Tummalacherla M, Snyder P, Bot BM, et al. Detecting the impact of subject characteristics on machine learning-based diagnostic applications. *npj Digit Med [Internet].* 2019;2(1):1–6. Available from: <http://dx.doi.org/10.1038/s41746-019-0178-x>

43. Audacity Development Team. Audacity [Internet]. 2018. Available from: <https://www.audacityteam.org/>
44. MathWorks. MATLAB. Natick, Massachusetts: The MathWorks Inc.; 2019.
45. Schultz BG. The Schultz MIDI Benchmarking Toolbox for MIDI interfaces, percussion pads, and sound cards. *Behav Res Methods*. 2019;51(1):204–34.
46. Lartillot O, Toivainen P, Eerola T. A matlab toolbox for music information retrieval. In: *Data analysis, machine learning and applications*. Springer; 2008. p. 261–8.
47. Tsanas A, Little MA, McSharry PE, Ramig LO. Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity. *J R Soc interface*. 2011;8(59):842–55.
48. Shue Y-L, Keating P, Vicens C, Yu K. Voicesauce. p Progr available online <http://www.seas.ucla.edu/spapl/voicesauce/> UCLA. 2009;
49. Hancock JT, Khoshgoftaar TM. CatBoost for big data: an interdisciplinary review. *J big Data*. 2020;7(1):1–45.

Methods

Participants

Healthy controls were recruited through advertisements within Australia. Clinical groups were recruited through medical centers in Australia. We recruited people diagnosed with multiple sclerosis ($N=112$) and Friedreich ataxia ($N=73$) as well as healthy controls ($N=229$). All patient participants were diagnosed by a physician and confirmed genetically for Friedreich ataxia. Demographic information of participant groups is shown in Table 2. Some participants were recorded on more than one occasion, leading to a larger final number of speech tokens per group: Multiple sclerosis ($N=787$), Friedreich ataxia ($N=158$), and healthy controls ($N=483$). To ensure that results were not driven by profiles of individuals, data were averaged over participants, resulting in one data point per participant for training and test phases (see (42)).

Apparatus

A condenser headset microphone (AKG C520, AKG Acoustic, Vienna, Austria) positioned 8-10cm from the mouth at an angle of 45° recorded speech. A Roland Quad-capture external soundcard connected to a Dell laptop using captured speech through Audacity software (43) and Redenlab® software at a sampling rate of 44.1kHz.

Table 2 *Demographic information for healthy controls (HC), Friedreich ataxia (FA), and multiple sclerosis (MS).*

Variable	Statistic	Group		
		HC	FA	MS
Age (years)	Mean	56.44	38.78	47.04
	SD	14.61	13.04	11.42
	Min	21	7	22
	Max	92	68	71
Sex	Female	126	39	83
	Male	104	41	29
Mean disease duration (years)	Mean	NA	22.89	13.66
	SD	NA	12.67	9.14
Mean severity score*	Mean	NA	79.67	3.36
	SD	NA	35.88	2.15

* FA = Friedreich ataxia rating scale (between 0 and 117), MS = Expanded disability status scale (between 0 and 10). These are not comparable measures of severity and are administered after diagnosis of the disease.

Procedure

Participants performed a DDK task where the syllables /PA/, /TA/, and /KA/ were repeated in an alternating fashion as many times as possible within one breath for a maximum of 10 seconds. Speech recordings were screened prior to feature analysis to manually remove speech artefacts and background noise.

Acoustic Feature Extraction

Acoustic features were extracted using custom-made MATLAB scripts that used standard signal processing functions from MATLAB (44), onset and offset detection algorithms (45), beat detection algorithms (25), music information retrieval (46), and speech analysis toolboxes (47,48). Acoustic features consisted of summary statistics (mean, standard deviation, coefficient of variation, minimum, maximum, range) of 74 variables that measure different aspects of speech qualia. These features include speech rate, utterance duration, pause duration, fundamental frequency, the first five formants, intensity, summed and peak energy across frequency bands, spectral decrease and spread, and a range of other spectral features used in the acoustic clinical marker literature (see supplementary materials for a full list and additional references).

Machine Learning

We used CatBoost as our machine learning classification algorithm. CatBoost is an open-source decision tree-based algorithm with gradient boosting and hardware optimisation (49). The main advantage of CatBoost over other algorithms is that it builds symmetric trees, employs weighted sampling, and performs ordered boosting. It also lowers the weights of variables that are less useful in identifying groups. These features decrease the need for hyperparameter tuning and reduces the chance of overfitting (49). Cross-validation was performed using 67%-33% Train-Test splits with 100 resamples using stratification to achieve the same balance for each class.

Figures

Figure 1

Mean, dispersion, and range for classification accuracy (A = F1 Accuracy, B = Precision, C = Recall) for healthy controls, and groups with Friedreich ataxia and multiple sclerosis for the full model, and models using a subset of acoustic features using the maximum group-wise average SHAP value cut-offs of 2% and 5%.

Figure 2

Receiver operating curves for healthy controls (HC), Friedreich ataxia (FA), and multiple sclerosis (MS) obtained from the model with average performance.

a

b

Figure 3

a: Normalized (z-scores) values of the top 10 acoustic features for identifying members of the healthy control (HC), Friedreich ataxia (FA), and multiple sclerosis (MS), groups. Note. See [here](#) for interactive figures for all acoustic features. b: Normalized (z-scores) values of the top 11 to 21 acoustic features for identifying members of the healthy control (HC), Friedreich ataxia (FA), and multiple sclerosis (MS) groups. Note. See [here](#) for interactive figures for all acoustic features.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SchultzVogelMLSupplementaryMaterials.docx](#)
- [SupplementaryMaterialsSHAPRankings.csv](#)