Vocal markers from sustained phonation in Huntington's Disease

Rachid Riad, Hadrien Titeux, Laurie Lemoine, Justine Montillot, Jennifer
Hamet Bagnou, Xuan Nga Cao, Emmanuel Dupoux, Anne-Catherine
Bachoud-Lévi

Vocal markers from sustained phonation in Huntington's Disease

Rachid Riad[1,2], Hadrien Titeux[1], Laurie Lemoine[2,3], Justine Montillot[2,3], Jennifer Hamet Bagnou[2,3], Xuan Nga Cao[1], Emmanuel Dupoux[1], Anne-Catherine Bachoud-Lévi[2,3]

[1] CoML/ENS/CNRS/EHESS/INRIA/PSL Research University, Paris, France
[2] NPI/ENS/INSERM/UPEC/PSL Research University, Paris, France
[3] Huntington's Disease National Reference Center, Neurology Department, Henri-Mondor Hospital, APHP, Créteil, France

`rachid.riad@ens.fr`

Abstract

Disease-modifying treatments are currently assessed in neurodegenerative diseases. Huntington's Disease represents a unique opportunity to design automatic sub-clinical markers, even in premanifest gene carriers. We investigated phonatory impairments as potential clinical markers and propose them for both diagnosis and gene carriers follow-up. We used two sets of features: Phonatory features and Modulation Power Spectrum Features. We found that phonation is not sufficient for the identification of sub-clinical disorders of premanifest gene carriers. According to our regression results, Phonatory features are suitable for the predictions of clinical performance in Huntington's Disease.

Index Terms: Huntington's disease, Phonation, Pathological speech processing, Dysarthria, Modulation Power Spectrum.

1. Introduction

Huntington's disease is an autosomal dominant neurodegenerative disease with a complete penetrance [1]. The disease is characterised by a triad of symptoms (motor, cognitive and psychiatric [2]) that leads to the progressive disability functioning of the individual with HD. The age of the clinical onset is highly variable (mean around 45 year-old) and the symptoms gradually worsen over 15 to 20 years until the death. The motor assessment allows defining a cut-off score that splits between premanifest and manifest gene carriers and thus the disease onset. Therefore, some of the deficits cognitive or psychiatric impairments or sub-clinical disorders may appear long before the disease manifests.

The automatic identification of pre-symptomatic carriers of the gene of Huntington's Disease and the prediction of the clinical scores are of particular interest for the neurological practice [3]. Indeed, having ecological markers of these clinical endpoints may: (1) help to prevent detrimental and harmful life events, (2) speed-up clinical trials, (3) increase the understanding of the disease in ground truth condition. Here, by using speech analysis of sustained phonation, we aim at distinguishing, in individuals, stage of the disease and predicting their clinical scores, as assessed by neurologists and neuropsychologists.

We rely on the fact that individuals with Huntington's Disease can exhibit different disorders during speech [4, 5] and language production [6, 7]. Motor speech disorders in Huntington's Disease are commonly referred to as *hyperkinetic* dysarthria [8]: variable rate, abnormal prosody, imprecise consonants and distorted vowels, phonation deviations, and sudden forced breath. Yet, there is some recent evidence that the speech motor impairments in Huntington's Disease are highly heterogeneous [9].

Here, we considered the speech features collected from simple recordings of the French vowel /a/ uttered in a sustained manner for as long as possible, in regular clinical conditions. We collected production of these vowels for Healthy Controls (C), gene carriers without overt manifestation of Huntington's Disease (preHD) and manifest gene carriers of Huntington's Disease (HD). We modelled each vowel with several features, which are impaired in hyperkinetic dysarthria. We investigated two sets of features: (1) Phonatory features already proven useful to distinguish gene carriers from control (preHD vs Controls [4] and HD vs Controls [10]), (2) Modulation Power Spectrum features to measure the modulations that characterise speech intelligibility [11] and roughness [12]. First, we conducted a statistical analysis at the group level of these features. Then, based on these sets of features, we used regularised linear models to predict the group and clinical scores of the patients.

Our statistical analyses and classification results showed that HD patients distinguished from controls, whereas the boundaries around preHD are more blurred. The Modulation Power Spectrum features complemented the Phonatory features to help identify preHD and improve the F1-score. In contrast, we observed that the Phonatory features have the best predictive capabilities of the clinical scores.

2. Related work

Previous studies have used sustained phonation for the assessment of various neurodegenerative disorders such as Parkinson's disease [13], Amyotrophic Lateral Sclerosis [14] and also Huntington's Disease [10]. These studies have extracted a number of hand-crafted features from the sustained phonation and built discriminative models to distinguish patients from controls. Yet, collecting pre-symptomatic speech data on other acquired neurodegenerative diseases is difficult, and can only be done retrospectively [15].

Rusz et al. [4] provided the closest study to ours using

Table 1: *Participants demographics and clinical scores*

	Controls	Huntington's disease Gene carriers	
Sub-groups	C	PreHD	HD
N	24	16	45
Gender	12F/12M	9F/7M	27F/18M
Age (years)	54.1 (8.9)	50.2 (12.2)	53.9 (11.3)
CAG Triplets	≤ 35	41.4 (1.5)	44.3 (3.4)
cUHDRS [16]	—	17.6 (1.4)	9.0 (3.6)
TFC [17]	—	13.0 (0.0)	10.3 (2.2)
TMS	—	0.38 (1.0)	36.0 (15.9)

sustained phonation for the automatic identification of preHD. They show great discriminative performance between preHD and controls.

Perez et al. [18] trained a Speech Recognition system on Huntington's speech and extracted a number of language features ranging from speech rate, and goodness of pronunciation to utterance length. The goal of this study differed from ours as they grouped together the preHD and HD patients in a single group to distinguish from control. They also pointed out to the difficulty to identify preHD.

We did not find any study attempting to predict the clinical scores in Huntington's disease from speech. Our strategy, which may allow following up individuals remotely, is likely to apply to other neurodegenerative diseases, such as in the Parkinson's Disease [19] or Alzheimer Disease [20].

3. Voice Database

Eighty five participants were included from two observational cohorts (NCT01412125 and NCT03119246) in this ancillary study at the Hospital Henri-Mondor Créteil, France): 61 people tested with a number of CAG repeats on the Huntingtin gene above 35 [1] (CAG > 35), and 24 Healthy Controls (C) (See Table 1). All participants signed an informed consent. Mutant Huntingtin gene carriers were considered premanifest if both they score less than five at the Total Motor score (TMS) and their Total functional capacity (TFC) equals 13 [21] using the Unified Huntington Disease Rating Scale (UHDRS) [22].

All participants completed a standardised speech battery. The data were annotated with Seshat [23] and Praat [24] softwares. The annotators were second-year graduate students in speech pathology, all French native speakers.

Each participant was asked to take a deep breath and to sustain the vowel /a/ at a constant intensity and pitch level for as long as possible. The recordings were done in the same condition for all participants, with a ZOOM H4n Pro recorder, sampled at 44.1 kHz with a 16-bit resolution.

Table 2: *List of Phonatory features based on [4]. SD stands for Standard Deviation.* †*The vocal tremor features could not be computed on all samples.*

Dimension	Features
Airflow insufficiency	Maximum Phonation Time
	First Occurrence of Voice Break
Aperiodicity	Number of Voice Breaks
	Degree of Pitch Breaks
	Degree of Vocal Arrests
Irregular vibration of vocal folds	F0 SD
	Recurrence Period Density Entropy
Signal perturbation	Jitter (local)
	Shimmer (local)
Increased noise	Harmonics to Noise Ratio
	Detrended Fluctuation Analysis
Vocal tremor†	Frequency Tremor Intensity Index
	Amplitude Tremor Intensity Index
Articulatory deficiency	Mean of SD of MFCC
	Mean of SD of Delta MFCC

4. Features

4.1. Phonatory Features

We used the Phonatory features from [10, 4] to measure dimensions that can impede the correct sustained production of the vowel /a/ for HD and preHD gene carriers: airflow insufficiency, aperiodicity, irregular vibration of vocal folds, signal perturbation, increased noise, vocal tremor, articulatory deficiency. These features are summarised in the Table 2.

To automatically extract these features, we used the Praat software [24], the Parselmouth wrapper [25] and the tremor package [26]. The fundamental frequency (F0) and MFCCs were obtained with the Kaldi toolkit [27]. Besides, we implemented the Detrended Fluctuation Analysis and the Recurrence Period Density Entropy features from [13], in Python. To compute these two features and replicate [13], we down-sampled the audio to 22.5 kHz, otherwise the audio is down-sampled to 16 kHz. Due to instability for the extraction of the tremor features, there are missing data points for this dimension (See Table 3).

To overcome the limitations of heterogeneity of acoustic methodologies across studies and libraries, we provide an open-source version of the code (link) to reproduce our results and to extract each dimension of the sustained phonation of the vowel /a/.

4.2. Modulation Power Spectrum features

To complete the Phonatory features, we used the Modulation Power Spectrum (MPS) extracted for each vowel /a/ of each individual. Different perceptual attributes occupy distinct areas of the MPS: roughness, gender and size characteristics of the speaker, and linguistic meaning [12]. Besides, the MPS captures the spectral and amplitude modulations of the pitch and its harmonics [1]. The MPS representation is the amplitude spectrum of the 2D Fourier Transform of a time-frequency representation obtained from the sound waveform [11, 28]. This time-frequency representation is a log-scaled amplitude of a spectrogram computed every 1 millisecond (ms) with a spacing between each frequency bin of 50Hz using a Gaussian window. The linear spacing in the frequency axis [11] better describes sounds that have harmonic structure, like long steady vowels. The MPS is computed every 10 ms with a 100 ms window of the spectrogram, then the final representation is averaged over the full duration of the sound. Upward and downward temporal modulations are kept between -200 Hz and +200 Hz, and spectral modulations between 0 and 9.5 Cycles/kHz (99% of the energy was found between these intervals).

5. Methods

5.1. Group level: Statistical analysis

We followed the process for statistical comparison and correction from Rusz et al., 2014 [4]. Given the non-normality of the data, we tested the differences between groups with the non-parametric Kruskal-Wallis test. The Bonferroni correction is applied to correct for the seven types of speech deficits (see Table 2). For post-hoc analyses, we applied the Kolmogorov-Smirnov test to all features to check for normality within each group, and the Levene's test for homoscedasticity between groups. If normality and homoscedasticity requirements were fulfilled, we applied an independent t-test otherwise; we applied a non-parametric equivalent (Mann-Whitney U-test). The p-values were Bonferroni corrected for multiple comparisons per feature. We also estimated the effect size with the Cohen's d. Results are summarised in Table 3. Statistical analyses for the MPS features are displayed in the arxiv version of this paper.

---

[1]Readers can refer to [11, 12] for in-depth study of the different perceptual areas

Table 3: *Results of the statistical analyses between the three groups for the sustain of the vowel /a/: Controls (C), asymptomatic genetic carrier of Huntington's Disease (preHD), symptomatic genetic carrier of Huntington's Disease (HD). The p-values significativity of the tests (Kruskal-Wallis tests H-statistic (H stat) and post-hoc pairwise tests on the Cohen's d) are reported with *: $0.01 < p \leq 0.05$, **: $0.001 < p \leq 0.01$, ***: $p \leq 0.001$. The post-hoc statistics are corrected for multiple comparison feature wise. SD stands for standard deviation. † 23 HD, 6 preHD, 6 Controls data points could not be computed for the Frequency Tremor Intensity Index. ‡ 15 HD, 6 preHD, 3 Controls data points could not be computed for the Amplitude Tremor Intensity Index.*

	Mean (SD)			H stat	Effect size Cohen's d		
	C	preHD	HD		HD/PreHD	HD/C	preHD/C
Phonatory Features							
Maximum Phonation Time (s)	17.0 (8.7)	15.7 (4.9)	9.1 (5.1)	23.6***	-1.27***	-1.19***	-0.18
First Occurrence of Voice Break (s)	13.2 (7.6)	10.3 (5.7)	6.3 (5.1)	16.6***	-0.77**	-1.14***	-0.41
Number of Voice Breaks	3.6 (9.2)	7.2 (12.3)	3.8 (7.4)	2.9	-0.37	0.02	0.34
Degree of Pitch Breaks (%)	1.1 (3.1)	4.8 (9.4)	6.3 (10.6)	5.7	0.15	0.59**	0.59
Degree of Vocal Arrests (%)	1.6 (2.8)	3.8 (5.9)	7.3 (10.1)	6.8	0.39	0.68**	0.49
F0 SD (Hz)	7.3 (10.8)	13.8 (15.2)	16.8 (15.3)	13.1**	0.20	0.68***	0.51
Recurrence Period Density Entropy	0.60 (0.18)	0.57 (0.16)	0.56 (0.16)	0.52	-0.09	-0.23	-0.13
Jitter (local) (%)	0.73 (0.47)	1.1 (1.7)	1.0 (1.0)	0.70	-0.10	0.32	0.36
Shimmer (local) (%)	8.5 (3.6)	7.9 (4.0)	8.4 (4.5)	0.37	0.14	0.02	0.15
Harmonics to Noise Ratio (dB)	16.1 (4.1)	15.8 (5.2)	15.3 (5.1)	0.52	-0.11	-0.19	-0.07
Detrended Fluctuation Analysis	0.70 (0.04)	0.70 (0.04)	0.68 (0.06)	2.5	-0.52	-0.41	0.16
Frequency Tremor Intensity Index †	4.0 (4.3)	7.1 (6.4)	7.4 (8.1)	3.5	0.07	0.52	0.60
Amplitude Tremor Intensity Index ‡	24.2 (7.6)	29.4 (14.7)	27.2 (14.4)	13.7**	0.57	1.10	0.50
Mean of SD of MFCC	6.6 (1.2)	6.6 (0.79)	8.5 (1.7)	25.6***	1.23***	1.21***	0.01
Mean of SD of Delta MFCC	1.3 (0.19)	1.2 (0.25)	1.5 (0.38)	14.7**	0.80**	0.77***	-0.13

5.2. Individual level: Machine Learning

To assess the performance of each set of features for classification and regression, we conducted 100 repeated learning-testing with 20% of the data left out as test set [29]. We used scikit-learn for all our models and data processing [30].

5.2.1. Group classification

We first compared the predictive power of the different input features to discriminate between the three groups: Controls (C), preHD, and HD patients. To do so, we trained a logistic regression regularised with ElasticNet (e.g. L1 and L2 combined with $C = 1$ and $ratio = 0.5$). Then, we computed the mean and standard deviation of the Accuracy and the F1 score (See Table 4). We also reported the chance level by selecting randomly the class based on the prior distribution.

5.2.2. Regression of the clinical scores

Second, we assessed the predictive capabilities of the features to predict the composite score cUHDRS [16] (currently used in international clinical trials [31]), the TFC and the TMS. We trained a Linear regression model a with ElasticNet regulariser (e.g. L1 and L2 combined with $C = 1$ and $ratio = 0.5$). The phonation from Controls are not used in the regression. Then, we computed the mean and standard deviation of the Mean Absolute Error and the coefficient of Determination R^2 (See Table 4) as well as the chance level by predicting the mean of each outcome on the train set.

6. Results and discussions

Statistical Analyses in Table 3 showed that the most affected dimensions in HD are those related to airflow insufficiency and articulatory deficiency. We did not find any significant difference between PreHD and Controls, while both groups differed from HD. Post-hoc analyses revealed that HD differed from from Controls but not from preHD for pitch/F0 related features:

the Degree of Pitch Breaks, the Degree of Vocal Arrests, and Standard deviation of the F0. This suggests an impairment of the vocal cords control prior to the clinical onset. None of the Phonatory features were sufficient to distinguish preHD from Controls. Tremor features displayed the strongest effect size when comparing preHD and Controls. But the unequal level of data loss in each group (51% HD, 38% preHD, and 25% Controls) suggests defining estimation methods that incorporate tremor instability and better tremor tracking methods [14] for future work.

The classification results (Table 4) for the Phonatory features yielded lower discrimination than the MPS features. Phonatory features performances are lower than the studies [4, 10]. This can be explained by the difference in the definition of the preHD group: the preHD population were based on a previous definition (the Diagnostic Confidence Level, item 17 of the UHDRS Motor Assessment). Some of their preHD genetic carrier showed some motor deficits (TMS than was up to 8 in their population). These differences may be due to the small sample size, inter-individual variability. The differences can also be attributed to the difficulty to tease apart the 3 groups in comparison to 2 groups comparison only. Besides, the MPS features are more suitable to identify the preHD. We also show the type of errors made by each set of features with the Confusion Matrices (see Figure 2). HD are the most identifiable group. Clearly, the preHD are often confused by the model as HD or Controls. Yet the types of errors differ. With the Phonatory features the preHD are even more classified as Controls than the Controls themselves (0.68 versus 0.45), which suggests a compensation mechanism.

The weights of the logistic regression trained to classify the sub-group based on the MPS features are interpretable. These weights can be visualised in the Figure 1. Even though the models had no prior how close the features are in the MPS space, we saw the emergence of patterns. The HD sub-classifier showed an area of activations for spectral modulation around 4 cycles/kHZ, which can be associated with temporal modula-

Figure 1: *Averaged weights of the Logistic Regression regularised with ElasticNet applied on the Modulation Power Spectrum Features to discriminate between each sub-group. Mean Sparsity = 37.1%*
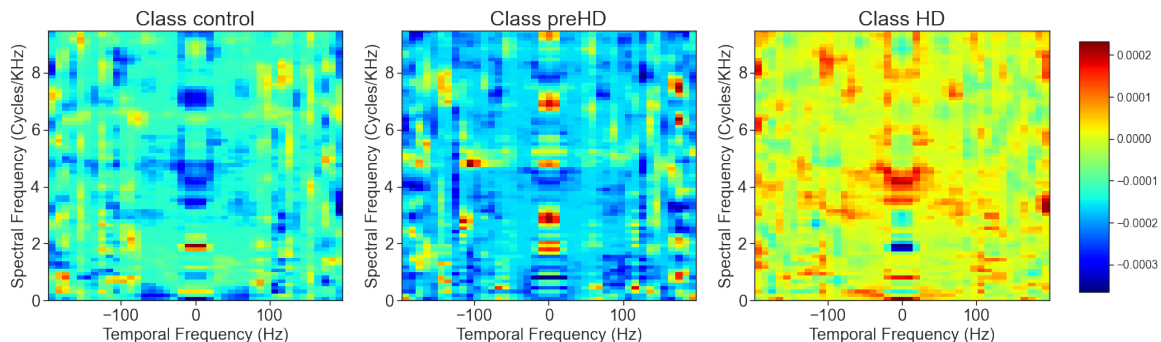

Table 4: *Results of the machine learning experiments based on the set of features obtained from the sustain of the vowel /a/. Mean and Standard Deviation are reported for each metrics. Classification performance is reported with Accuracy and F1-macro score. Regression performance is reported with the Mean Absolute Error (MAE) and the Coefficient of Determination (R^2). Best score for each metric is reported in **bold**. † The Phonatory features do not include the tremor features as they could not be computed for each subject.*

| | Classification Results | | Regression Results | | | | | |
| | C vs preHD vs HD | | cUHDRS | | TFC | | TMS | |
	Accuracy	F1-macro	MAE	R^2	MAE	R^2	MAE	R^2
Random Prior	0.38 (0.1)	0.31 (0.1)	4.21 (0.6)	0.0 (0.0)	1.96 (0.3)	0.0 (0.0)	19.07 (2.3)	0.0 (0.0)
Phonatory feat.†	**0.56 (0.1)**	0.40 (0.1)	**2.78 (0.5)**	**0.58 (0.2)**	**1.64 (0.3)**	**0.37 (0.2)**	**13.14 (1.7)**	**0.53 (0.2)**
MPS	0.54 (0.1)	0.43 (0.1)	3.68 (0.5)	0.28 (0.1)	1.83 (0.3)	0.15 (0.1)	17.87 (3.2)	0.26 (0.2)
MPS+Phonatory feat.†	**0.56 (0.1)**	**0.46 (0.1)**	3.10 (0.5)	0.49 (0.1)	1.70 (0.3)	0.25 (0.1)	15.34 (3.1)	0.29 (0.2)

Figure 2: *Confusion matrices for the Logistic regression based on the Phonatory features (left) and on the MPS features (right) averaged across all the repeated learning-testing experiments.*
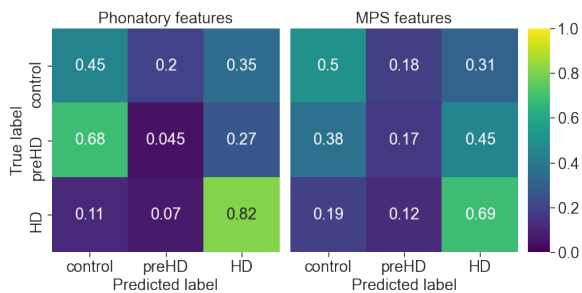
tions between -45 and 45 Hz. We found the strongest activation at the origin (0.0, 0.0), which relates to voice breaks during the phonation. Even though the classifier is not perfect for preHD, we saw several specific activations along the Spectral modulation at 2.5 Cycles/kHZ and 7 Cycles/kHZ at Temporal Frequency equal to 0. This might suggest Frequency modulations specific to preHD, trying to avoid the zone around 4 Cycles/kHz of HD. The combination of the set of MPS and Phonatory features improved the classification performances up to an Accuracy of 0.56 and F1-macro of 0.46.

In contrast, Phonatory features better reflected the clinical scores cUHDRS, TFC, and TMS (Table 4) The composite measure cUHDRS, currently used in the assessment of international clinical trial [31], is the best predicted among the scores, if we rank them based on the coefficient of determination R^2. This means that the Phonatory features are a better indicator of the severity of the disease, once the clinical onset is declared.

7. Conclusions

Here, we combine the data from three groups for the study of the vocal markers of sustained phonation in Huntington's Disease patients: symptomatic, pre-symptomatic and control. We applied a statistical analysis, a classification study and assessed the capabilities to predict clinical scores. In addition, we introduced Modulation Power Spectrum features, to more traditional Phonatory features. Airflow insufficiency and articulatory deficiency measures distinguished HD patients from both preHD and Controls. However, Modulation Power Spectrum features provided more hope of distinguishing preHD from Controls. They allowed a three-fold reduction in misidentification of preHD. When replicated in a larger scale and in another population, this suggest that speech phonation might replace long traditional assessments, considering that the the sustained vowel task takes less than 1 minute and UHDRS takes a minimum of 30 minutes when ran by experts. It may allow repetitive testing with limited retest effect and recordings could also be blindly scored and analysed. This points to speech as a major future tool in the clinical panel of assessments.

8. Acknowledgements

9. References

[1] J. F. Gusella, N. S. Wexler, P. M. Conneally, S. L. Naylor, M. A. Anderson, R. E. Tanzi, P. C. Watkins, K. Ottina, M. R. Wallace, A. Y. Sakaguchi *et al.*, "A polymorphic dna marker genetically linked to huntington's disease," *Nature*, vol. 306, no. 5940, pp. 234–238, 1983.

[2] M. J. Novak and S. J. Tabrizi, "Huntington's disease," *Bmj*, vol. 340, p. c3109, 2010.

[3] E. J. Wild and S. J. Tabrizi, "One decade ago, one decade ahead in huntington's disease," *Movement Disorders*, vol. 34, no. 10, pp. 1434–1439, 2019.

[4] J. Rusz, C. Saft, U. Schlegel, R. Hoffman, and S. Skodda, "Phonatory dysfunction as a preclinical symptom of huntington disease," *PloS one*, vol. 9, no. 11, 2014.

[5] A. P. Vogel, C. Shirbin, A. J. Churchyard, and J. C. Stout, "Speech acoustic markers of early stage and prodromal huntington's disease: a marker of disease onset?" *Neuropsychologia*, vol. 50, no. 14, pp. 3273–3278, 2012.

[6] W. Hinzen, J. Rosselló, C. Morey, E. Camara, C. Garcia-Gorro, R. Salvador, and R. de Diego-Balaguer, "A systematic linguistic profile of spontaneous narrative speech in pre-symptomatic and early stage huntington's disease," *Cortex*, vol. 100, pp. 71–83, 2018.

[7] M. Teichmann, E. Dupoux, S. Kouider, P. Brugières, M.-F. Boissé, S. Baudic, P. Cesaro, M. Peschanski, and A.-C. Bachoud-Lévi, "The role of the striatum in rule application: the model of huntington's disease at early stage," *Brain*, vol. 128, no. 5, pp. 1155–1167, 2005.

[8] F. L. Darley, A. E. Aronson, and J. R. Brown, "Differential diagnostic patterns of dysarthria," *Journal of speech and hearing research*, vol. 12, no. 2, pp. 246–269, 1969.

[9] S. K. Diehl, A. S. Mefford, Y.-C. Lin, J. Sellers, K. E. McDonell, M. de Riesthal, and D. O. Claassen, "Motor speech patterns in huntington disease," *Neurology*, vol. 93, no. 22, pp. e2042–e2052, 2019.

[10] J. Rusz, J. Klempır, E. Baborová, T. Tykalová, V. Majerová *et al.*, "Objective acoustic quantification of phonatory dysfunction in huntington's disease." *PloS one*, vol. 8, no. 6, pp. e65 881–e65 881, 2013.

[11] T. M. Elliott and F. E. Theunissen, "The modulation transfer function for speech intelligibility," *PLoS computational biology*, vol. 5, no. 3, 2009.

[12] L. H. Arnal, A. Flinker, A. Kleinschmidt, A.-L. Giraud, and D. Poeppel, "Human screams occupy a privileged niche in the communication soundscape," *Current Biology*, vol. 25, no. 15, pp. 2051–2056, 2015.

[13] M. A. Little, P. E. McSharry, S. J. Roberts, D. A. Costello, and I. M. Moroz, "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection," *Biomedical engineering online*, vol. 6, no. 1, p. 23, 2007.

[14] J. Peplinski, V. Berisha, J. Liss, S. Hahn, J. Shefner, S. Rutkove, K. Qi, and K. Shelton, "Objective assessment of vocal tremor," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 6386–6390.

[15] V. Berisha, J. Liss, T. Huston, A. Wisler, Y. Jiao, and J. Eig, "Float like a butterfly sting like a bee: Changes in speech preceded parkinsonism diagnosis for muhammad ali," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, vol. 2017, 2017, pp. 1809–1813.

[16] S. A. Schobel, G. Palermo, P. Auinger, J. D. Long, S. Ma, O. S. Khwaja, D. Trundell, M. Cudkowicz, S. Hersch, C. Sampaio *et al.*, "Motor, cognitive, and functional declines contribute to a single progressive factor in early hd," *Neurology*, vol. 89, no. 24, pp. 2495–2502, 2017.

[17] I. Shoulson, "Huntington disease: functional capacities in patients treated with neuroleptic and antidepressant drugs," *Neurology*, vol. 31, no. 10, pp. 1333–1333, 1981.

[18] M. Perez, W. Jin, D. Le, N. Carlozzi, P. Dayalu, A. Roberts, and E. M. Provost, "Classification of huntington disease using acoustic and lexical features," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, vol. 2018, 2018, pp. 1898–1902.

[19] K. M. Smith, J. R. Williamson, and T. F. Quatieri, "Vocal markers of motor, cognitive, and depressive symptoms in parkinson's disease," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 2017, pp. 71–78.

[20] S. Luz, F. Haider, S. de la Fuente, D. Fromm, and B. MacWhinney, "Alzheimer's dementia recognition through spontaneous speech: The ADReSS Challenge," in *Proceedings of INTERSPEECH 2020*, Shanghai, China, 2020. [Online]. Available: https://arxiv.org/abs/2004.06833

[21] S. J. Tabrizi, D. R. Langbehn, B. R. Leavitt, R. A. Roos, A. Durr, D. Craufurd, C. Kennard, S. L. Hicks, N. C. Fox, R. I. Scahill *et al.*, "Biological and clinical manifestations of huntington's disease in the longitudinal track-hd study: cross-sectional analysis of baseline data," *The Lancet Neurology*, vol. 8, no. 9, pp. 791–801, 2009.

[22] K. Kieburtz, J. B. Penney, P. Corno, N. Ranen, I. Shoulson, A. Feigin, D. Abwender, J. T. Greenarnyre, D. Higgins, F. J. Marshall *et al.*, "Unified huntington's disease rating scale: reliability and consistency," *Neurology*, vol. 11, no. 2, pp. 136–142, 2001.

[23] H. Titeux*, R. Riad*, X.-N. Cao, N. Hamilakis, K. Madden, A. Cristia, A.-C. Bachoud-Lévi, and E. Dupoux, "Seshat: A tool for managing and verifying annotation campaigns of audio data," in *Twelfth International Conference on Language Resources and Evaluation (LREC 2020)*, Marseille, May 2020, * Equal contribution.

[24] P. Boersma and D. Weenink, "Praat, a system for doing phonetics by computer," *Glot international*, vol. 5, 2002.

[25] Y. Jadoul, B. Thompson, and B. de Boer, "Introducing Parselmouth: A Python interface to Praat," *Journal of Phonetics*, vol. 71, pp. 1–15, 2018.

[26] M. Brückl, "Vocal tremor measurement based on autocorrelation of contours," in *Thirteenth Annual Conference of the International Speech Communication Association*, 2012.

[27] P. Ghahremani, B. BabaAli, D. Povey, K. Riedhammer, J. Trmal, and S. Khudanpur, "A pitch extraction algorithm tuned for automatic speech recognition," in *2014 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2014, pp. 2494–2498.

[28] J. E. Elie and F. E. Theunissen, "Zebra finches identify individuals using vocal signatures unique to each call type," *Nature communications*, vol. 9, no. 1, pp. 1–11, 2018.

[29] G. Varoquaux, P. R. Raamana, D. A. Engemann, A. Hoyos-Idrobo, Y. Schwartz, and B. Thirion, "Assessing and tuning brain decoders: cross-validation, caveats, and guidelines," *NeuroImage*, vol. 145, pp. 166–179, 2017.

[30] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[31] S. J. Tabrizi, B. R. Leavitt, G. B. Landwehrmeyer, E. J. Wild, C. Saft, R. A. Barker, N. F. Blair, D. Craufurd, J. Priller, H. Rickards *et al.*, "Targeting huntingtin expression in patients with huntington's disease," *New England Journal of Medicine*, vol. 380, no. 24, pp. 2307–2316, 2019.

Figure 3: *Results of the statistical analyses for the Modulation Power Spectrum Features between the three groups for the sustain of the vowel /a/: Controls (C), asymptomatic genetic carrier of Huntington's Disease (preHD), symptomatic genetic carrier of Huntington's Disease (HD). Left figure is the distribution of the H-statistic. Middle figure is the distribution of the uncorrected p-values. Right figure is the distribution of the FDR corrected p-values.*
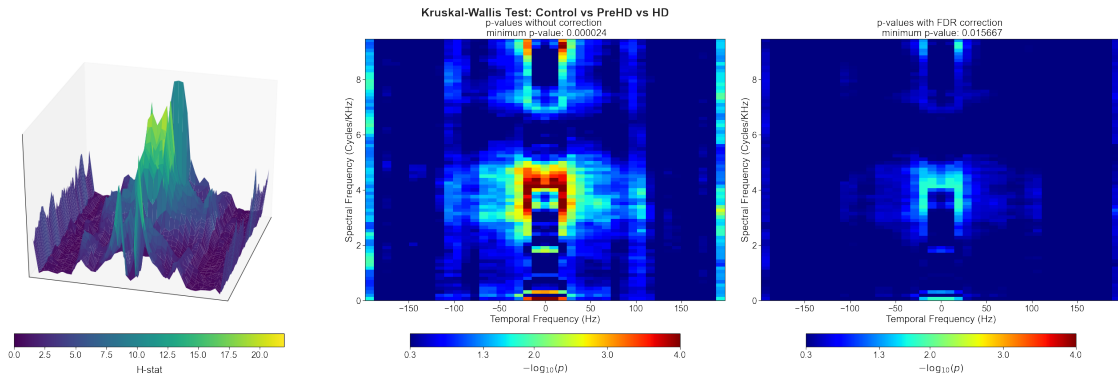
10. Appendix

10.1. Additional Statistical analysis for the Modulation Power Spectrum Features

We tested the differences between groups with the non-parametric Kruskal-Wallis test. The False-Discovery-Rate correction is applied to correct for the multiple comparisons ($77 \times 41 = 3157$). We reported the statistical results in the Figure 3. 1.7% (14.6%) of the MPS features are found significant with (without) the FDR correction. We found significant area to separate the three groups of activation, especially around 4 cycles/kHz (very similar to the area found to identify the symptomatic HD patients).

10.2. Interpretation of coefficients of linear models for the regression of clinical scores based on the Phonatory features

As we used linear models, each target value is modelled as a linear combination of the input features. We followed the methodology analysis from the example of scikit-learn [30] (link). The stability of the predictors is shown through the different coefficients across folds. As we used the ElasticNet Regulariser (L1+L2 regularisation of the coefficients), we also observed the selection of variables based on the Coefficient Importance analysis. The results for the regression results for the clinical measures cUHDRS, TFC, TMS are reported respectively in the Figure 4, Figure 5, Figure 6.

Figure 4: *Coefficient Importance of the different Phonatory Features across the different cross-validation folds to predict the cUHDRS*
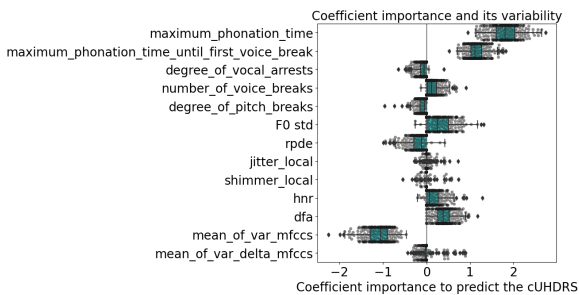
The Maximum Phonation Time and First Occurrence of Voice Break are the features the most used for all 3 regression

Figure 5: *Coefficient Importance of the different Phonatory Features across the different cross-validation folds to predict the TFC*
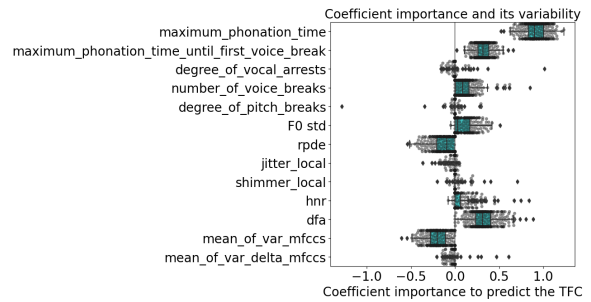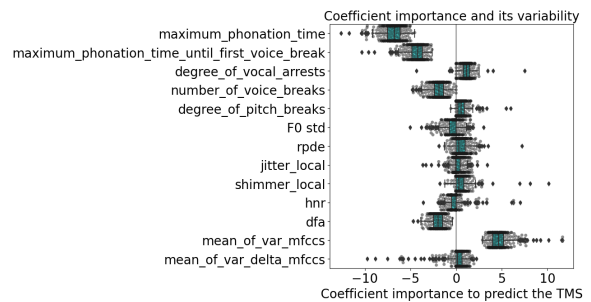

Figure 6: *Coefficient Importance of the different Phonatory Features across the different cross-validation folds to predict the TMS*

tasks. The Mean of SD of MFCC also contribute to the prediction of the cUHDRS and TMS. The Detrended Fluctuation Analysis is also a feature useful for the prediction of the TMS (coefficient never set to 0).

Otherwise, the contributions of the other features for the other tasks are more blurred or often set to 0. These coefficient importance give also the direction associated with the progress of the scores and then the disease.