ASHA American Speech-Language-Hearing Association

AJSLP

Research Article

Language-Independent Acoustic Biomarkers for Quantifying Speech Impairment in Huntington's Disease

Vitória S. Fahed,[a,b] Emer P. Doheny,[a,b] Carla Collazo,[c] Joanna Krzysztofik,[d] Elliot Mann,[e] Philippa Morgan-Jones,[e,f] Laura Mills,[e] Cheney Drew,[e] Anne E. Rosser,[g] Rebecca Cousins,[h] Grzegorz Witkowski,[d] Esther Cubo,[c] Monica Busse,[e] and Madeleine M. Lowery[a,b]

[a] School of Electrical and Electronic Engineering, University College Dublin, Ireland [b] Insight Centre for Data Analytics, University College Dublin, Ireland [c] Hospital Universitario de Burgos, Spain [d] Institute of Psychiatry and Neurology, Warsaw, Poland [e] Centre for Trials Research, Cardiff University, United Kingdom [f] School of Engineering, Cardiff University, United Kingdom [g] Brain Repair Centre and BRAIN Unit, Schools of Medicine and Biosciences, Cardiff University, United Kingdom [h] North Bristol NHS Trust, United Kingdom

ARTICLE INFO

Article History:
Received May 18, 2023
Revision received September 8, 2023
Accepted January 14, 2024

Editor-in-Chief: Rita R. Patel
Editor: Panying Rong

https://doi.org/10.1044/2024_AJSLP-23-00175

ABSTRACT

Purpose: Changes in voice and speech are characteristic symptoms of Huntington's disease (HD). Objective methods for quantifying speech impairment that can be used across languages could facilitate assessment of disease progression and intervention strategies. The aim of this study was to analyze acoustic features to identify language-independent features that could be used to quantify speech dysfunction in English-, Spanish-, and Polish-speaking participants with HD.
Method: Ninety participants with HD and 83 control participants performed sustained vowel, syllable repetition, and reading passage tasks recorded with previously validated methods using mobile devices. Language-independent features that differed between HD and controls were identified. Principal component analysis (PCA) and unsupervised clustering were applied to the language-independent features of the HD data set to identify subgroups within the HD data.
Results: Forty-six language-independent acoustic features that were significantly different between control participants and participants with HD were identified. Following dimensionality reduction using PCA, four speech clusters were identified in the HD data set. Unified Huntington's Disease Rating Scale (UHDRS) total motor score, total functional capacity, and composite UHDRS were significantly different for pairwise comparisons of subgroups. The percentage of HD participants with higher dysarthria score and disease stage also increased across clusters.
Conclusion: The results support the application of acoustic features to objectively quantify speech impairment and disease severity in HD in multilanguage studies.
Supplemental Material: https://doi.org/10.23641/asha.25447171

Huntington's disease (HD) is a rare, inherited neurogenerative disorder characterized by motor, cognitive, and behavioral dysfunction. Changes in voice and speech are among the characteristic symptoms of HD associated with disease progression. Speech can be severely impaired in HD and is characterized as hyperkinetic dysarthria, which is associated with neurodegeneration within the basal ganglia (Duffy, 2005; Hamilton et al., 2012). Dysarthria in HD can manifest as changes in pitch and prosody, and irregular speech patterns and speech timing, affected by deficits in respiration, phonation, and articulation secondary to disordered muscle activity and control (Rusz, Saft, et al., 2014; Skodda et al., 2014; Vogel et al., 2012).

Speech assessment is currently performed by neurologists as part of the Unified Huntington's Disease Rating Scale (UHDRS) using a dysarthria subcomponent as part of the motor assessment section (Huntington Study Group, 1996). Speech impairment is also evaluated by speech-language pathologists using perceptual measures to assess phonation, articulation, prosody, resonance, and

Correspondence to Vitória S. Fahed: vitoria.fahed@ucdconnect.ie.
Disclosure: *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

control of respiratory function and orofacial movements (Behrman, 2021; Duffy, 2005; Hamilton et al., 2012), and using standardized assessments such as the Sentence Intelligibility Test (SIT; Yorkston et al., 1996). While the UHDRS is the gold standard for overall disease monitoring in HD and perceptual measures are the gold standard for speech impairment and dysarthria evaluation in clinical diagnosis, monitoring, and disease management, they are subject to interrater variability, and scoring can be granular, often with low resolution (Clinch et al., 2018; Duffy, 2005). Objective, quantitative measures of speech function that can be performed consistently within the clinic or home and translated into clinical metrics or person-centered outcomes offer the potential for monitoring disease progression and guiding therapeutic interventions. Ideally, such methods should be language-independent to enable comparison of patients across different clinical sites or languages. This is particularly important in rare diseases such as HD where multicenter investigations are necessary to reach the patient numbers required for clinical trials. Language has been shown to affect acoustic features and phoneme representations within the brain (Näätänen et al., 1997; Styler, 2017). Although multilanguage approaches have been investigated for speech in other conditions, such as Parkinson's disease (Orozco-Arroyave et al., 2016), at present, language-independent methods have not been investigated in HD populations. Furthermore, while objective and quantitative evaluation methods for speech have been developed for research purposes (Boersma & Weenink, 2022; Boyce et al., 2012), their translatability to clinical applications remains a challenge. There remains a need for objective, efficient, clinically relevant methods to quantify speech impairment in HD that are suitable for both clinical and remote evaluation that are easily deployed and simply administered by nonspecialists (Vogel et al., 2012). The use of mobile technologies can support routine deployment of quantitative methods for recording and analyzing acoustic features in the clinic or at home. Used appropriately, mobile devices can provide a reliable method of speech recording for analysis of acoustic and temporal features, provided that the device type and position are appropriately controlled (Fahed et al., 2022).

Acoustic and temporal features provide a means with which to quantitatively assess speech in HD. A number of quantitative speech studies in HD to date have been conducted in a premanifest HD cohort with the aim of developing speech biomarkers or predicting the onset of speech symptoms (Kaploun et al., 2011; Rusz, Saft, et al., 2014; Vogel et al., 2012, 2016). Studies conducted in manifest HD population analyzing both acoustic and temporal features have been limited by sample size, application to a single language, or only one type of speech task (Kouba et al., 2023; Rusz et al., 2013; Rusz, Klempíř, et al., 2014; Skodda et al., 2014). Among patients with manifest HD, an improved understanding of changes in acoustic and temporal features may inform clinical decision making and provide the basis for objective markers of disease stage and progression. Clinical scoring systems provide easily interpreted measures of speech impairment at the cost of low resolution and high interrater variability. On the other hand, acoustic features yield a high-dimensional feature space that can be difficult to interpret clinically. Methods such as unsupervised cluster analysis can be used to identify subtypes within heterogeneous data and have been used in the analysis of sustained vowel recordings from individuals with Parkinson's disease (Tsanas & Arora, 2021) and rater-scored reading tasks in HD to reveal clusters within the data sets (Diehl et al., 2019; Kim et al., 2022).

Automatic quantification of acoustic features present an opportunity for development of speech biomarkers that can be used for characterization and monitoring of speech impairment in HD to complement clinical scoring systems and to aid speech-language pathologists by providing detailed information on changes in speech production in HD. In addition to providing information on physiological and pathogenic processes (FDA-NIH Biomarker Working Group, 2016), an important consideration in the development of such biomarkers is the robustness, specificity, generalizability, and clinical interpretability of the measures (Ramanarayanan et al., 2022).

Toward this goal, this study examined acoustic and temporal features in English-, Spanish-, and Polish-speaking participants with manifest HD, with the aim of identifying-language-independent biomarkers of speech impairment for disease monitoring in HD. Acoustic and temporal features from HD participants and age-, biological sex–, and language-matched control participants were first compared. A subset of features that differed between patient and control groups, but not across languages, was then identified. Principal component analysis (PCA) was applied to the language-independent features in the HD data to reduce the dimensionality of the data set, followed by unsupervised clustering to identify patient subgroups. Finally, the relationships between the identified clusters and clinical measures of speech impairment, disease stage, and motor symptoms were examined, revealing differences and a progressive increase in the level of clinical impairment across clusters identified from the speech features.

Materials and Method

Participants

Ninety participants with genetically confirmed HD disease (50.25 ± 12.72 years; 47 women, all manifest HD: 17 Stage 2, 68 Stage 3, five disease stages not available)

and 83 adults with no known history of neurodegenerative disease or speech disorder (47.18 ± 11.82 years; 45 women) gave their written consent to participate in the study. Participants were recruited from three clinical sites in Wales, Poland, and Spain, (see Table 1). The study was approved by the NHS research ethics committee (Wales REC 3, United Kingdom, 19/WA/0329, January 13, 2020) Ethical Committee for Research With Medicines of the Burgos and Soria Health Area, Junta Castilla and León (CElm 2296, Burgos, Spain, 1788-19//19/WA/0329, June 5, 2020), and Bioethics Committee at the Institute of Psychiatry and Neurology in Warsaw (Warsaw, Poland, 09-02-957, September 9, 2020), as part of the consortium project Multi-Domain Lifestyle Targets for Improving Prognosis in Huntington's Disease. Participants with HD were recruited from participants in the Enroll-HD study, and their clinical data were made available by the CHDI Foundation, Inc. Enroll-HD is a global clinical research platform designed to facilitate clinical research in HD. Core data sets are collected annually from all research participants as part of this multicenter longitudinal observational study. Data are monitored for quality and accuracy using a risk-based monitoring approach.

Acoustic data were recorded in the clinic during sustained vowel phonation (SV), syllable repetition (SR), and reading passage (RP) tasks in the participant's native language using a mobile device (Samsung Galaxy Tab A6 in Wales and Poland; Huawei Mate 10 lite and Samsung A51 in Spain). Participants were seated in a comfortable position in a quiet room, with the device placed in front of them on a desk at a distance of approximately 5 cm from the participant's torso (see Supplemental Material S1).

Participants completed the following three sequential and alternating motor rate tasks: SV ([a:]; Patel et al., 2018; Rusz et al., 2021), RP in their native language (English: Rainbow Passage [Fairbanks, 1960; Patel et al., 2018], Spanish: Brief passage about doctors [Orozco-Arroyave et al., 2016], Polish: North Wind and the Sun [Pettorino et al., 2017]), and SR (Rusz et al., 2021; [pa] SR1, [ta] SR2, [ka] SR3, [pataka] SR4, and [pati] SR5). Participants were asked to sustain the vowel "ah" for as long as they could during one breath. They were instructed to read the passage at their own pace, as they would naturally read out loud. Finally, they were asked to repeat each syllable as fast and as clearly as they could for 5 s or until task failure. Each task was repeated 2 or 3 times. Audio data were sampled at 44.1 or 48 kHz, 16-bit analogue-to-digital conversion, and saved in an uncompressed .wav format using the Android application Easy Voice Recorder. This application was chosen since it provides a function to access the unprocessed data from the microphone, which was the setting selected. One participant from the Spanish group was excluded in the analysis as they did not perform all syllables in the SR test.

Demographic and clinical e-information were also recorded. Clinical data consisted of the total motor score and dysarthria score from the UHDRS, total functional capacity, the disease burden score (Penney et al., 1997), and the composite UHDRS (cUHDRS; Schobel et al., 2017). A voluntary movement score was also calculated from the sum of the finger-tapping, tongue protrusion, pronation, and supination of the forearm, luria, and bradykinesia UHDRS components. Similarly, an involuntary movement score was calculated as the sum of the rigidity, chorea, and dystonia from the UHDRS. Saccade and gait were not included in the voluntary and involuntary movement scores.

Signal Analysis

Preprocessing and Voiced/Unvoiced Signal Detection

The audio-recorded signals were filtered with a zero-phase eighth-order Butterworth bandpass filter between 10 Hz and 5 kHz. All signals were then down-sampled to 44.1 kHz, the mean was removed, and 0.5 s of data at the

Table 1. Demographic distribution of control participants and participants with Huntington's disease (HD).

Language	Group	Number of participants	Age ± *SD* (years)	Disease stage	CAG repeats	Composite UHDRS	Disease burden score
English	Control	24 (10 females)	46.54 ± 10.54	—	—	—	—
English	HD	29 (9 females)	50.41 ± 12.05	Stage 2: 6; Stage 3: 22; not available: 1	43.54 ± 3.76	10.08 ± 4.37	376.14 ± 127.05
Spanish	Control	36 (18 females)	48.72 ± 12.05	—	—	—	—
Spanish	HD	37 (20 females)	51.84 ± 12.41	Stage 2: 11; Stage 3: 23; not available: 3	43 ± 2.63	10.93 ± 5.19	368.6 ± 101.81
Polish	Control	23 (17 females)	45.44 ± 12.42	—	—	—	—
Polish	HD	24 (18 females)	47.63 ± 14.03	Stage 3: 23; not available: 1	45.58 ± 12.8	8.8 ± 3.73	422.31 ± 88.45

Note. UHDRS = Unified Huntington's Disease Rating Scale; CAG = Cytosine, Adenine, Guanine; — = does not apply.

start and end of each signal was discarded, which represented silent periods before and after the speech test started. Finally, the signal was padded with zeros for 2 s at the start and end, and each signal was normalized with respect to its maximum amplitude. A voiced/unvoiced detector based on the Teager–Kaiser energy operator (Kaiser, 1990) was used to detect the voiced part of each signal using the methodology described by Fahed et al. (2022).

Feature Estimation

Acoustic and temporal features were estimated as previously described (Fahed et al., 2022) and summarized in Table 2. For the SV, a 0.5-s moving window with 50% overlap was used. Features were estimated for each window and averaged across windows. The entire signal between each detected onset/offset was used for feature estimation. The acoustic features were grouped into frequency-based, amplitude-based, and cepstral-based features (see Table 2) and estimated for each speech task (SV, RP, SR).

Temporal features were also estimated for each task (see Table 2). For SV, maximum phonation time, first occurrence of a voice break, degree of vocal arrests, and number of voice breaks were estimated. For RP and SR, total speech time, pause ratio, net speech rate, and interval duration from one vocalization to another (intDur) were estimated. The maximum syllable repetition capacity (maxSylRep) was also estimated for the SR test.

All features were estimated using custom Python scripts, with the exception of shimmer and its variants for which the amplitude series, composed of the glottal cycles, was extracted using the DYPSA-VOICEBOX library in MATLAB (Brookes, 1997). The control data and Python code for estimating acoustic features can be made available upon request.

Statistical Analysis

Linear mixed models (Winter, 2013) were implemented in R (Bates et al., 2014) to investigate differences in the features across each cohort (controls and participants with HD) for each speech task separately. For all features, age, biological sex, and language were included as fixed effects, and the intercept for participants was included as a random effect. For the SR task, all five syllables were included in the same model, and syllable was added a random slope. p values were obtained by comparing the model with a null model without the effect of condition using analysis of variance (ANOVA). The Benjamini–Hochberg false discovery rate (FDR) was used as a post hoc test to account for multiple comparisons within each speech task (SV, RP, and SR). Values less of .05 for the corrected p value were considered significant. The effect of language on features in the participants with HD was examined using linear mixed models to identify features on which language had a significant effect. Multiple comparisons were corrected using the FDR.

Feature Transformation and Clustering

Following statistical analysis to examine the differences between participants with HD and control participants, the acoustic and temporal features, which were significantly different between HD and control groups but which did not differ across languages, were identified. PCA (Hotelling, 1933) was then applied to these features to reduce the dimensions of the feature set for further analysis in the HD population. To account for effects of biological sex, all features were standardized separately for the "female" and "male" categories. The principal components (PCs) that accounted for over 70% of the variance of the selected features were selected. These PCs were then used to identify groups or clusters within the participants with HD using a k-means clustering algorithm.

The Python scikit-learn library (Pedregosa et al., 2011) was chosen with its standard configuration for k-means, which defines the cluster centroids based on empirical probability distribution. Four clusters were selected using both elbow method (Satopää et al., 2011) and by plotting the Silhouette scores for two to six clusters. The clustering process was evaluated using Silhouette (Rousseeuw, 1987), Caliński (Caliński & Harabasz, 1974), and Davies (Davies & Bouldin, 1979) scores, (see Figure 1).

Analysis of Clinical Data and Motor Impairment

Correlation analysis was performed between each feature and PC, with significant strong correlations ($\rho > 0.65$, $p < .05$) considered to indicate a significant contribution of a given feature toward a given PC. If the threshold ($\rho > 0.65$, $p < .05$) was not reached for any feature, the feature with the highest significant correlation coefficient was considered to have the greatest contribution to that PC.

The relationships between each identified cluster and clinical measures—UHDRS total motor score, total functional capacity, dysarthria, voluntary and involuntary movement scores, disease burden score, and cUHDRS—were then examined and compared between clusters using one-way ANOVA, followed by the Fisher's least significance differences for post hoc comparisons. The distribution of dysarthria scores, disease stage, and language across clusters was also examined. Finally, the distribution of language within each disease stage was examined to check for potential imbalances across groups.

Results

Differences in Acoustic and Temporal Features Between Participants With HD and Control Participants

Linear mixed models revealed significant differences in all features between control and HD groups during the SV task, with the exception of fundamental frequency,

Table 2. Acoustic and temporal features, their description, mathematical notation, and physiological meaning.

Group	Symbols	Features	Description	Equation[a]	Physiological equivalence[b]
Frequency-based acoustic features	*F*0 (Hz)	Fundamental frequency	Peak of the autocorrelation function	*F*0	Frequency of vocal fold vibration
	*SD F*0 (Hz)	Standard deviation of *F*0	Variation of the fundamental frequency	$\sigma = \sqrt{\frac{\Sigma (x_i-\mu)^2}{N}}$	Irregular vibration of vocal cords
	HNR (dB)	Harmonics-to-noise ratio	Ratio between the tonal components and noise in the voice	$HNR = 10 \times \log_2 \left(\frac{h}{1-h}\right)$, h = height autocorrelation function	Quantity of noise due to incomplete vocal fold closure
	Jitter (%)	Jitter	Variability of the glottal periods relative to two neighboring sample points	$Jitter = \frac{\Sigma_{i=2}^{N} \lvert T_i - T_{i-1} \rvert / (N-1)}{\Sigma_{i=1}^{N} T_i / N}$	Perturbation of the *F*0, variation of the voice range
	RAP (%)	Relative average perturbation of jitter	Average absolute difference between a sample and its difference between two neighboring samples, normalized to the mean glottal period	$RAP = \frac{\Sigma_{i=2}^{N-1} \lvert T_i - (T_{i-1}+T_i+T_{i+1})/3 \rvert / (N-2)}{\Sigma_{i=1}^{N} T_i / N}$	Analogous to jitter, can be a measure of voice quality and stability of the vocal fold vibrations
	PPQ5 (%)	5-point period perturbation quotient	Analogous to RAP, but considering four closest neighbors	$PPQ5 = \frac{\Sigma_{i=3}^{N-2} \lvert T_i - (T_{i-2}+T_{i-1}+T_i+T_{i+1}+T_{i+2})/5 \rvert / (N-4)}{\Sigma_{i=1}^{N} T_i / N}$	
	DDP (%)	Difference of differences of period	Average absolute second-order difference of a sample normalized to the mean glottal period	$DDP = \frac{\Sigma_{i=2}^{N-1} \lvert \ (T_{i+1}-T_i) - (T_i - T_{i-1}) \ \rvert / (N-2)}{\Sigma_{i=1}^{N} T_i / N}$	
Amplitude-based acoustic features	Shimmer (dB)	Shimmer	Variability of amplitude relative to two neighboring sample points	$Shimmer = \frac{\Sigma_{i=2}^{N} \lvert A_i - A_{i-1} \rvert / (N-1)}{\Sigma_{i=1}^{N} A_i / N}$	Perturbation of voice amplitude, variation of expiratory flow
	APQ3 (%)	3-point amplitude perturbation quotient	Average absolute difference between a sample and its difference between two neighboring samples, normalized to the mean amplitude	$APQ3 = \frac{\Sigma_{i=2}^{N-1} \lvert A_i - (A_{i-1}+A_i+A_{i+1})/3 \rvert / (N-2)}{\Sigma_{i=1}^{N} A_i / N}$	Analogous to shimmer, can be a measure of voice quality and stability of the expiratory flow
	APQ5 (%)	5-point amplitude perturbation quotient	Analogous to APQ3, but considering four closest neighbors	$APQ5 = \frac{\Sigma_{i=3}^{N-2} \lvert A_i - (A_{i-2}+A_{i-1}+A_i+A_{i+1}+A_{i+2})/5 \rvert / (N-4)}{\Sigma_{i=1}^{N} A_i / N}$	
	APQ11 (%)	11-point amplitude perturbation quotient	Analogous to APQ3, but considering five closest neighbors	$APQ11 = \frac{\Sigma_{i=12}^{N} \left\lvert \Sigma_{j=i-p}^{i} A_j - \frac{1}{N} \Sigma_{i=1}^{N} A_i \right\rvert / (N-10)}{\Sigma_{i=1}^{N} A_i / N}$	
	DDA (%)	Difference of differences of amplitude	Average absolute second order difference of a sample normalized to the mean amplitude	$DDA = \frac{\Sigma_{i=2}^{N-1} \lvert \ (A_{i+1}-A_i) - (A_i - A) \ \rvert / (N-2)}{\Sigma_{i=1}^{N} A_i / N}$	
	GNE (–)	Glottal-to-noise excitation ratio	Correlation of different frequency bands, that can reveal uncorrelated excitation	*GNE*	Presence of turbulent noise in the vibrations of the vocal tract

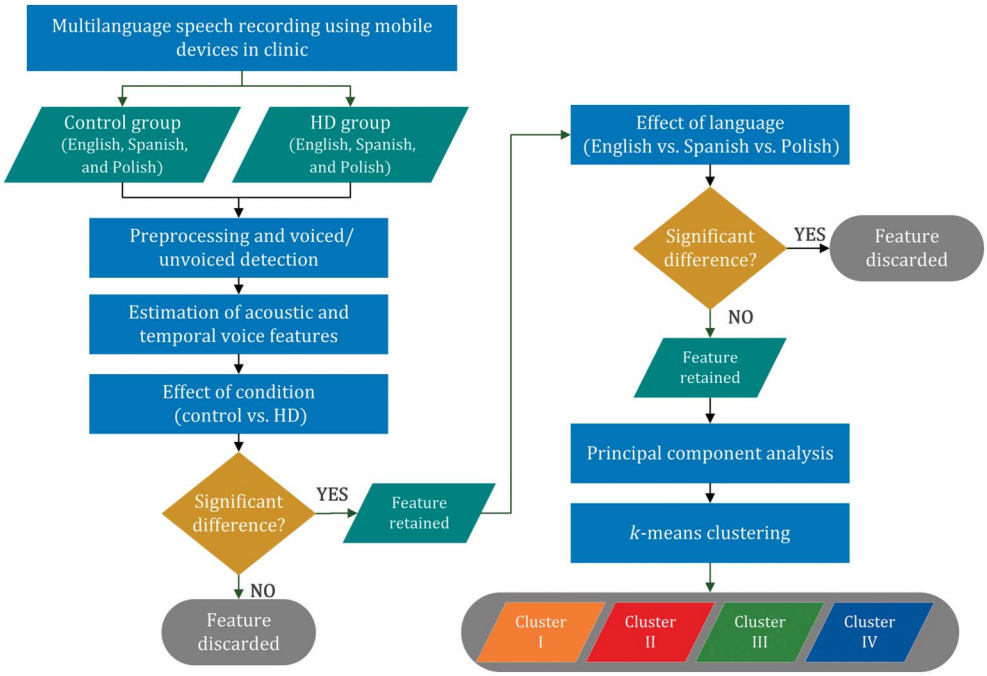
(*table continues*)

Table 2. *(Continued).*

Group	Symbols	Features	Description	Equation[a]	Physiological equivalence[b]
Cepstral-based acoustic features	*SD* MFCC (–)	Standard deviation of the Mel frequency cepstral coefficients	Variation of the individual MFCCs, which represent partition in the frequency domain	$\sigma_{MFCC} = \sqrt{\frac{\Sigma\,(MFCC_i - \mu_{MFCC})^2}{N_{MFCC}}}$	Reflect instability of vocal tract elements responsible for subtle motion of articulators
	SD Delta (–)	Standard deviation of the MFCC delta	Variation of the derivative of MFCCs	$\sigma_{\Delta} = \sqrt{\frac{\Sigma\,(\Delta_i - \mu_{\Delta})^2}{N_{\Delta}}}$	Analogous to *SD* MFCC, can reflect instability of vocal tract
	SD Delta 2 (–)	Standard deviation of the second order MFCC delta	Variation of second order derivative of MFCCs	$\sigma_{\Delta\Delta} = \sqrt{\frac{\Sigma\,(\Delta\Delta_i - \mu_{\Delta\Delta})^2}{N_{\Delta\Delta}}}$	
Temporal features	MPT (s)	Maximum phonation time	Duration of the longest sustained vowel if voice breaks were present	$MPT = onset_n - offset_n$ $n \rightarrow$pair with largest phonation	Lung capacity and airflow insufficiency
	FOVB (s)	First occurrence of voice break	Time stamp of the first voice break during vowel phonation	$FOVB = offset_0$	
	DVA (%)	Degree of vocal arrest	Total pause time over total phonation time	$DVA = \frac{\Sigma_{i=0}^{N}\, length(pause_i)}{\Sigma_{i=0}^{N}\, length(phonation_i)}$	Aperiodicity
	NVB (–)	Number of voice break	Number of voice breaks followed by restart	$NVB = onsets$	
	TST (s)	Total speech time	Time from the onset the offset of the last syllable	$TST = onset_0 - offset_n$	Slowness of speech
	PR (%)	Pause rate	Total pause time over total speech time	$PR = \frac{\Sigma_{i=0}^{N}\, length(pause_i)}{\Sigma_{i=0}^{N}\, length(speech_i)}$	Pace stability
	NSR (s)	Net speech rate	Total speech time minus total pause time	$NSR = \Sigma_{i=0}^{N}\, length(speech_i) - \Sigma_{i=0}^{N}\, length(pause_i)$	
	intDur (s)	Interval duration	Time (s) from the onset of one vocalization to the following	$intDur = \frac{\Sigma_{i=0}^{N-1}\, onset_{i+1} - onset_i}{N}$	
	maxSylRep (Hz)	Maximum syllable repetitions	Syllable repetition rate estimated over 5 s	$maxSylRep = \frac{onsets}{onset_0 - offset_n}$	Rhythm of repeated vocalization

[a]Information from Skodda et al., 2010, 2014. [b]Information from Boersma & Weenink, 2022; Rusz, Saft, et al., 2014; Skodda et al., 2010, 2014; Tsanas et al., 2011; Boersma, 1993.

Figure 1. Flowchart illustrating the data analysis process. HD = Huntington's disease.



glottal-to-noise excitation ratio (GNE), and Mel frequency cepstral coefficient (MFCC) delta (see Figures 2 and 3). During the RP task, significant differences between groups were also observed for the harmonics-to-noise ratio (HNR), 5-point amplitude perturbation quotient (APQ5), 11-point amplitude perturbation quotient (APQ11), standard deviation of the MFCC (*SD* MFCC), standard deviation of the MFCC delta (*SD* Delta), standard deviation of the second order MFCC delta (*SD* Delta 2; see Figure 2), and all temporal feature (see Figure 3). Similarly, during SR, significant differences were observed for the HNR, jitter, relative average perturbation of jitter (RAP), APQ5, APQ11, GNE, *SD* MFCC, *SD* Delta, *SD* Delta 2 (see Figure 2), and pause ratio, net speech rate, intDur, and maxSylRep (see Figure 3).

In addition to the effect of condition, a significant effect of language in the control population was observed on GNE and cepstral features in the SV task, all features from the RP task except *SD* fundamental frequency (*F*0), and a range of features during SR task. All features apart from *F*0, GNE, and cepstral features were selected from the SV task; no feature was selected from the RP task. From SR1 (/pa/) task, jitter, RAP, 5-point period perturbation quotient (PPQ5), difference of differences of period (DDP), *SD* MFCC, *SD* Delta 2, pause ratio, net speech rate, intDur, and maxSylRep were selected. From SR2 (/ta/), *SD* Delta 2, net speech rate, intDur, and maxSylRep were selected. From SR3 (/ka/), jitter, RAP, PPQ5, DDP, *SD* MFCC, *SD* Delta, *SD* Delta 2, net speech rate, intDur, and maxSylRep were selected. From SR4 (/pataka/), pause ratio, net speech rate, and intDur were selected. No feature was selected from SR5.

PCA Feature Transformation and Clustering

From the 46 identified language-independent features, the first five PCs, which accounted for more than 70% of the variance in the HD data, were selected. Each PC was correlated with the acoustic features estimated and used to build the feature space. The features exhibiting the highest correlation with each PC are presented in Table 3. The descriptive statistics for the PCs within each cluster of participants (see Table 4) reveal that Cluster I has the lowest overall mean for PC1, whereas Cluster II has the lowest mean for PC3 and PC4 and Cluster III has the lowest mean for PC2 and Cluster IV for PC5. Cluster II has the highest mean for PC5, whereas Cluster III has the highest mean for PC3 and Cluster IV has the highest mean for PC1, PC2, and PC4.

Relationship Between Speech Clusters and Clinical Scores

Clinical measures of motor impairment, disease burden, and cUHDRS increased from Cluster I to Cluster IV (see Figure 4). ANOVA, following post hoc analysis, confirmed significant differences between Clusters I and II (cUHDRS: 0.002), Clusters I and III (UHDRS total motor score: p = .007; total functional capacity: p = .005; voluntary movement score: p = .007; involuntary movement score: p = .17; disease burden score: p = .005; cUHDRS: p < .001), Clusters I and IV (UHDRS total motor score: p = .007; total functional

Figure 2. Acoustic features estimated for controls and participants with Huntington's disease (HD) during sustained vowel (a–i), syllable repetition (j–r), and reading passage (s–aa). *p < .05 corrected for multiple comparisons. *SD F*0 = standard deviation of fundamental frequency; HNR = harmonics-to-noise ratio; GNE = glottal-to-noise excitation ratio; *SD* MFCC = standard deviation of the Mel frequency cepstral coefficients; *SD* Delta = standard deviation of the MFCC delta; *SD* Delta 2 = standard deviation of the second-order MFCC delta.

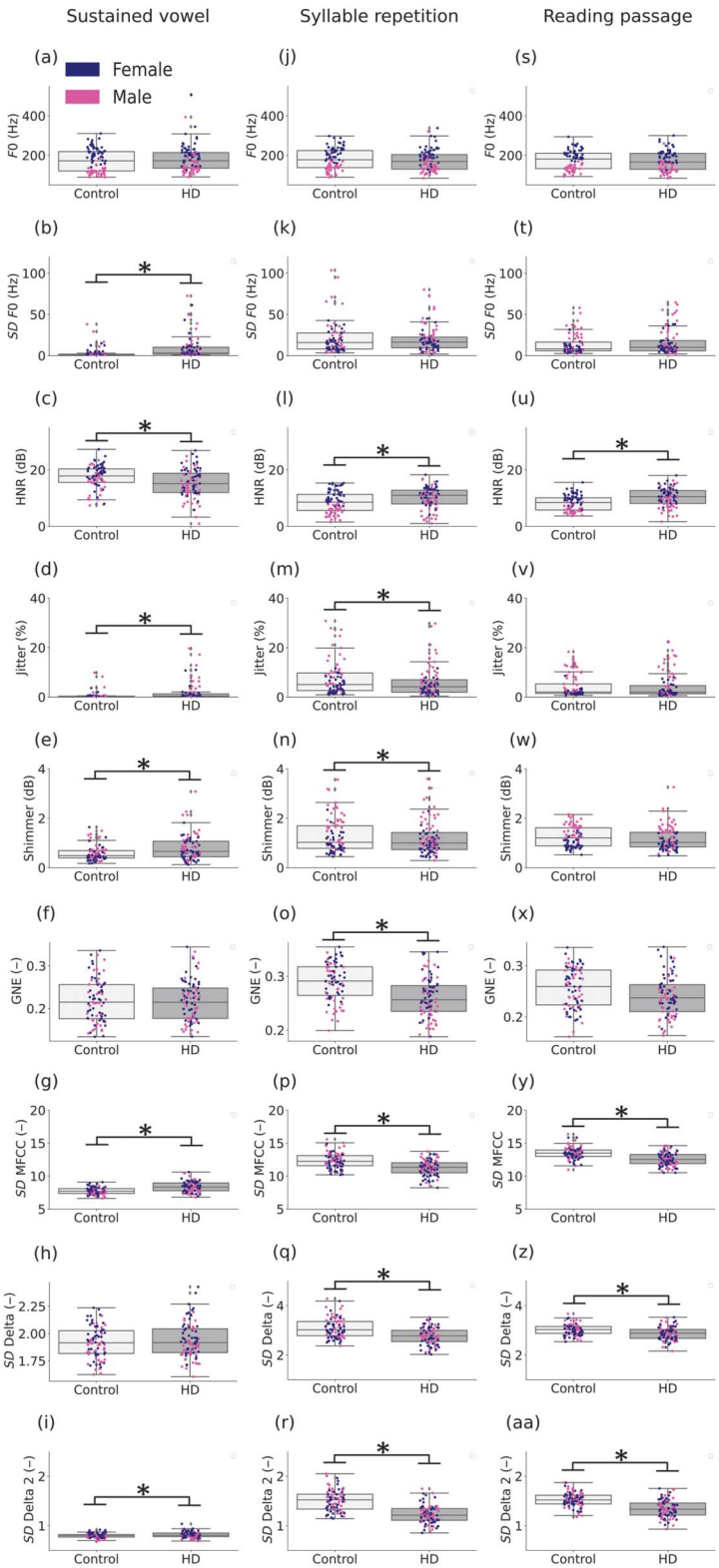
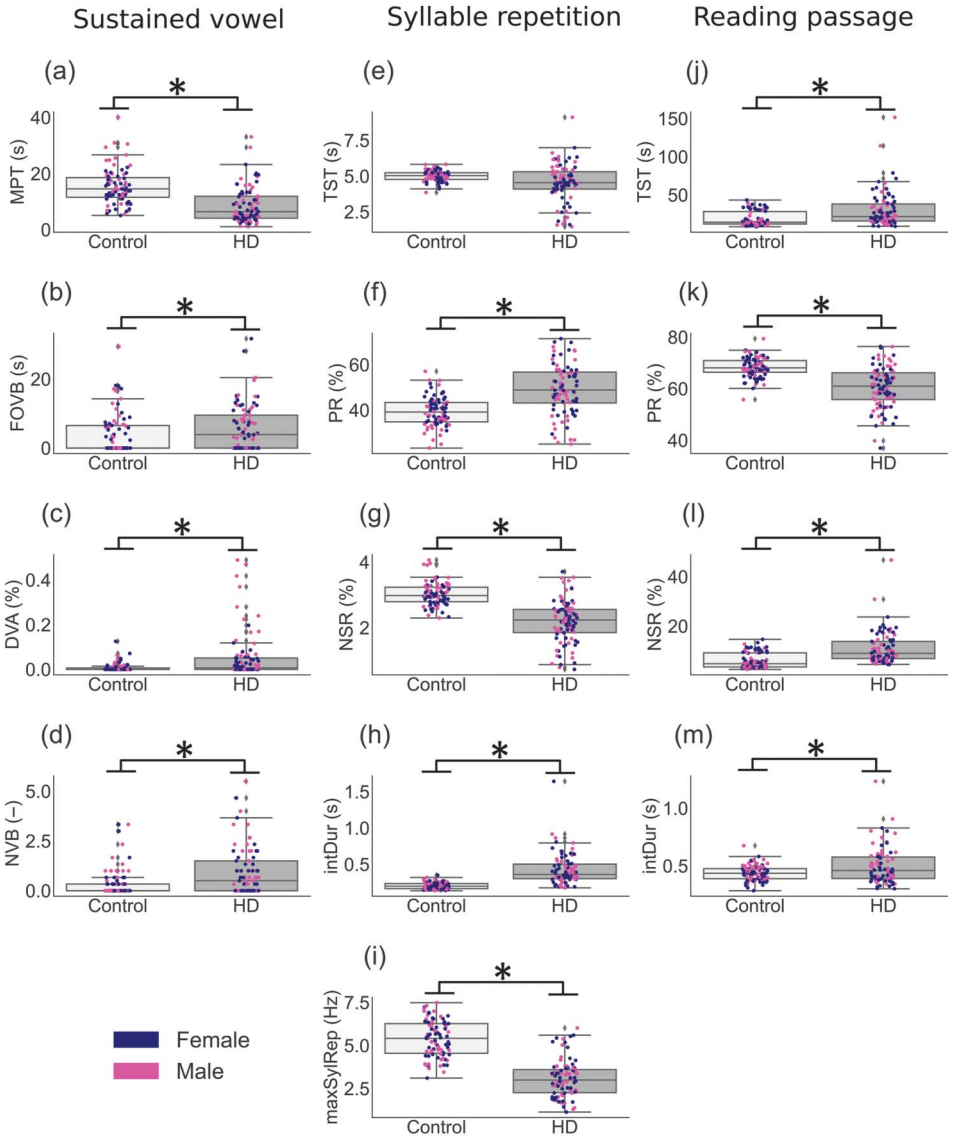


Figure 3. Temporal features estimated for controls and participants with Huntington's disease (HD) during sustained vowel (a–d), syllable repetition (e–i), and syllable repetition (j–m) tasks. *p < .05 corrected for multiple comparisons. MPT = maximum phonation time; FOVB = first occurrence of voice break; DVA = degree of vocal arrest; NVB = number of voice break; TST = total speech time; PR = pause rate; NSR = net speech rate; maxSylRep = maximum syllable repetition capacity; intDur = interval duration.



capacity: p = .012; voluntary movement score: p = .017; involuntary movement score: p = .024; cUHDRS: p = .002), and Clusters II and III (disease burden score: p = .032).

A high percentage of participants with dysarthria scores of 0, reflecting little or no impairment, fell into Clusters I and II (see Figure 5). Between 20% and 30% of patients within Clusters I and II had dysarthria scores of 1, and less than 5% of those in Cluster I had a score of 2, the highest dysarthria score within this cohort. Clusters III and IV exhibited a more mixed distribution, with approximately 40% of patients with scores of 0 or 1 and less than 20% with dysarthria scores of 2. Similar to the dysarthria scores, when evaluating disease stage, there is a clear trend between clusters (see Figure 5). Clusters I, II, and III contained patients in both disease Stages 2 and 3, whereas Cluster IV contained patients in disease Stage 3 only.

Discussion

Quantitative analysis of speech impairment using mobile technology has the potential to enable fully automated assessment of speech in the clinic and at home to provide objective biomarkers of disease monitoring and for assessing therapeutic interventions in HD. As many of

Table 3. Correlation between acoustic voice features and principal components extracted from principal component analysis ($p < .001$ for all).

Principal components	Speech test	Features	ρ
PC1	SV	HNR	−0.71
		Jitter	0.69
		RAP	0.72
		PPQ5	0.70
		DDP	0.72
		Shimmer	0.74
		APQ3	0.70
		APQ5	0.72
		APQ11	0.75
		DDA	0.70
	SR1	intDur	0.77
		maxSylRep	−0.78
	SR2	intDur	0.72
		maxSylRep	−0.73
	SR3	intDur	0.74
		maxSylRep	−0.74
PC2	SR1	Jitter	0.72
		PPQ5	0.71
	SR2	Jitter	0.73
		RAP	0.75
		PPQ5	0.71
		DDP	0.74
PC3	SR1	NSR	−0.85
	SR2	NSR	−0.82
	SR3	NSR	−0.83
PC4	SR1	*SD* MFCC	−0.55
PC5	SR3	DDP	0.57

Note. PC = principal component; SV = sustained vowel phonation; HNR = harmonics-to-noise ratio; RAP = relative average perturbation of jitter; PPQ5 = 5-point period perturbation quotient; DDP = difference of differences of period; APQ3 = 3-point amplitude perturbation quotient; APQ5 = 5-point amplitude perturbation quotient; APQ11 = 11-point amplitude perturbation quotient; DDA = difference of differences of amplitude; SR = syllable repetition; intDur = interval duration; maxSylRep = maximum syllable repetition capacity; NSR = net speech rate; *SD* MFCC = standard deviation of the Mel frequency cepstral coefficients.

these features vary across languages, to facilitate large-scale clinical trials, identification of speech features that are altered in HD and which can be assessed across languages is required. Here, we estimated acoustic and temporal features from speech recordings made with mobile devices from control participants and participants with manifest HD among native English, Spanish, and Polish speakers. Language independent features that differed between the control and HD groups were identified and examined using cluster analysis, which revealed subgroups with varying levels of speech impairment within the HD data set. The study represents the first multilingual study of acoustic speech features for disease monitoring in HD.

Significant differences were observed between HD and control participants across a range of acoustic and temporal features (see Figures 2 and 3). For the SV test, maximum phonation time, HNR, and *SD* MFCC were significantly lower in HD participants when compared with controls. Number of voice breaks, *SD* *F*0, jitter, and shimmer were found to be higher in HD. Mobile assessment findings are consistent with previous laboratory findings that reported lower maximum phonation time (Rusz et al., 2013; Skodda et al., 2014); lower HNR (Rusz et al., 2013); and higher number of voice breaks, *SD* *F*0, jitter, and shimmer (Rusz et al., 2013) in participants with HD during SV along with higher *SD* MFCC (Rusz et al., 2013).

During the RP task, total speech time, intDur, and *SD* *F*0 were all higher in the HD group. Pause ratio was lower, and no difference in *F*0 across groups was observed. This is also in agreement with previous findings that have observed higher total speech time (Skodda et al., 2014) and intDur during RP task for patients with HD (Hartelius et al., 2003; Skodda et al., 2014). It is in partial agreement with a study during a reading task and monologue that reported higher *F*0 in HD in males during the monologue and decreased *F*0 in females, though *SD* *F*0 was not significantly different in either task (Rusz, Klempíř, et al., 2014). The differences between these results and those reported here may be due to differences in tasks and the larger cohort in the present study, which may reveal differences not detected in a smaller population. In Hartelius et al. (2003), it was reported that *F*0 was increased in HD and *SD* *F*0 was significantly smaller, in contrast with our findings. *SD* *F*0 indicates the variation of the pitch during speech and can indicate irregular vocal fold vibrations, which can be affected in HD due to dysarthria.

In the SR task, significant differences were similarly observed between the patient and control groups in the majority of acoustic and temporal features examined. The observation of lower rate of SR (maxSylRep) for participants with HD contrasts with a previous study in 21 patients, which reported that maxSylRep did not differ between HD and control groups (Skodda et al., 2014), though was correlated with motor performance.

Analysis of the effect of language revealed differences for several of the features examined. An effect of language was observed on all of the HD-sensitive features from the RP task; the majority of the features in the SR tasks; and *F*0, GNE, and *SD* Delta in the SV task. An effect of language on acoustic features is well established in control groups and patients with Parkinson's disease (Kováč et al., 2022; Orozco-Arroyave et al., 2016). When comparing the phonology of English, Spanish, and Polish, all three languages are nontonal with a trochaic rhythm type; however, some Spanish dialects do not exhibit

Table 4. Mean and standard deviation (*M* ± *SD*) values of principal components in each cluster.

Clusters	PC1	PC2	PC3	PC4	PC5
I	–3.44 ± 1.85	1.91 ± 1.98	–0.33 ± 1.94	–0.56 ± 1.71	0.15 ± 1.89
II	3.02 ± 2.35	–0.88 ± 1.65	–1.60 ± 2.52	–0.75 ± 1.96	0.88 ± 1.20
III	–0.44 ± 1.62	–2.11 ± 1.26	1.13 ± 2.38	0.65 ± 1.61	–0.44 ± 0.90
IV	4.86 ± 3.90	8.38 ± 4.00	0.67 ± 1.43	1.22 ± 3.17	–1.31 ± 2.53

Note. PC = principal component.

rhythmic stress (Goedemans & van der Hulst, 2013; Maddieson, 2013b). English and Polish have a complex syllable structure, whereas Spanish has a moderately complex one (Maddieson, 2013a). The phonological differences in rhythm and syllable structure may partially account for the differences observed in the acoustic features of SR and RP across languages, while the fact that all three languages are nontonal may contribute to the similarity across languages observed for the SV task. Following the removal of those features that were found to be sensitive to language, 49 of the initial features were retained for feature reduction and clustering in the HD group.

Figure 4. (a) Unified Huntington's Disease Rating Scale (UHDRS) total motor score (TMS), (b) total functional capacity (TFC), (c) voluntary movement score (VMS), (d) involuntary movement score (IMS), (e) disease burden score (DBS), and (f) composite UHDRS (cUHDRS) for each of the identified HD clusters. Cluster I corresponds to the lowest levels of motor and speech impairment, and Cluster IV corresponds to the highest. *p < .05 corrected for multiple comparisons. HD = Huntington's disease.

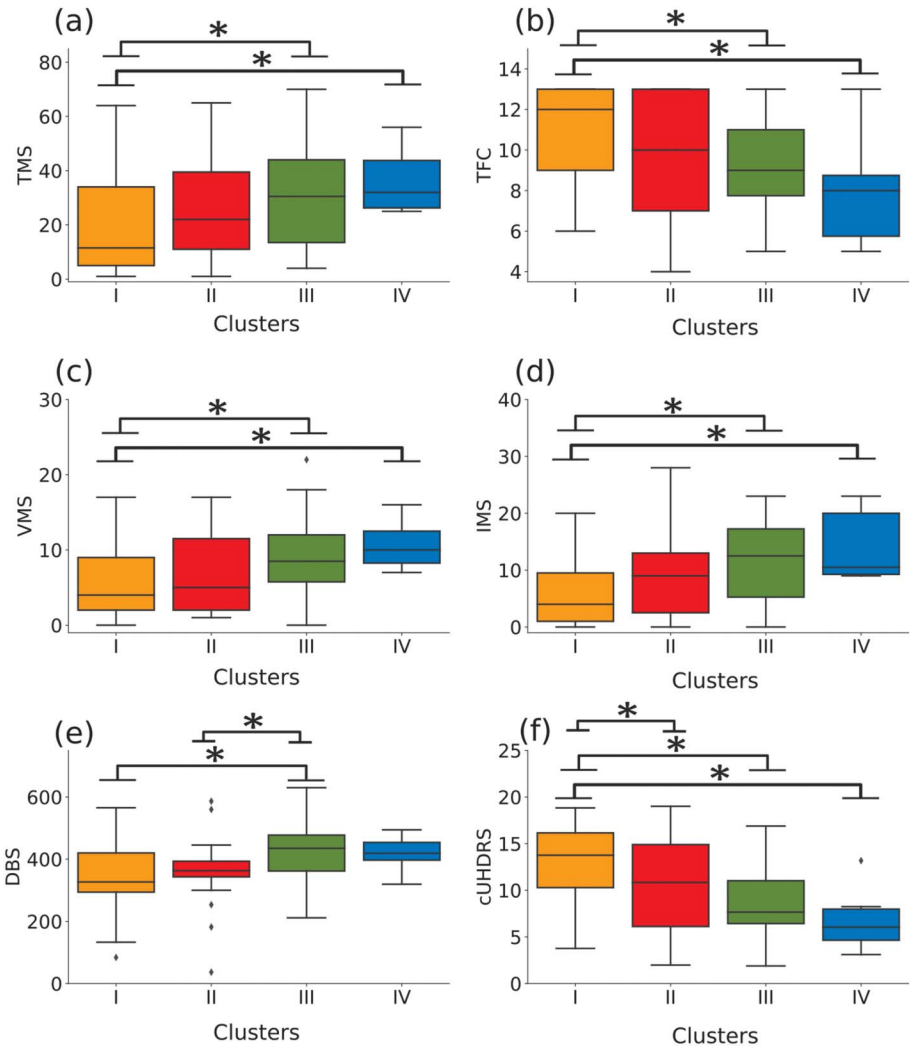
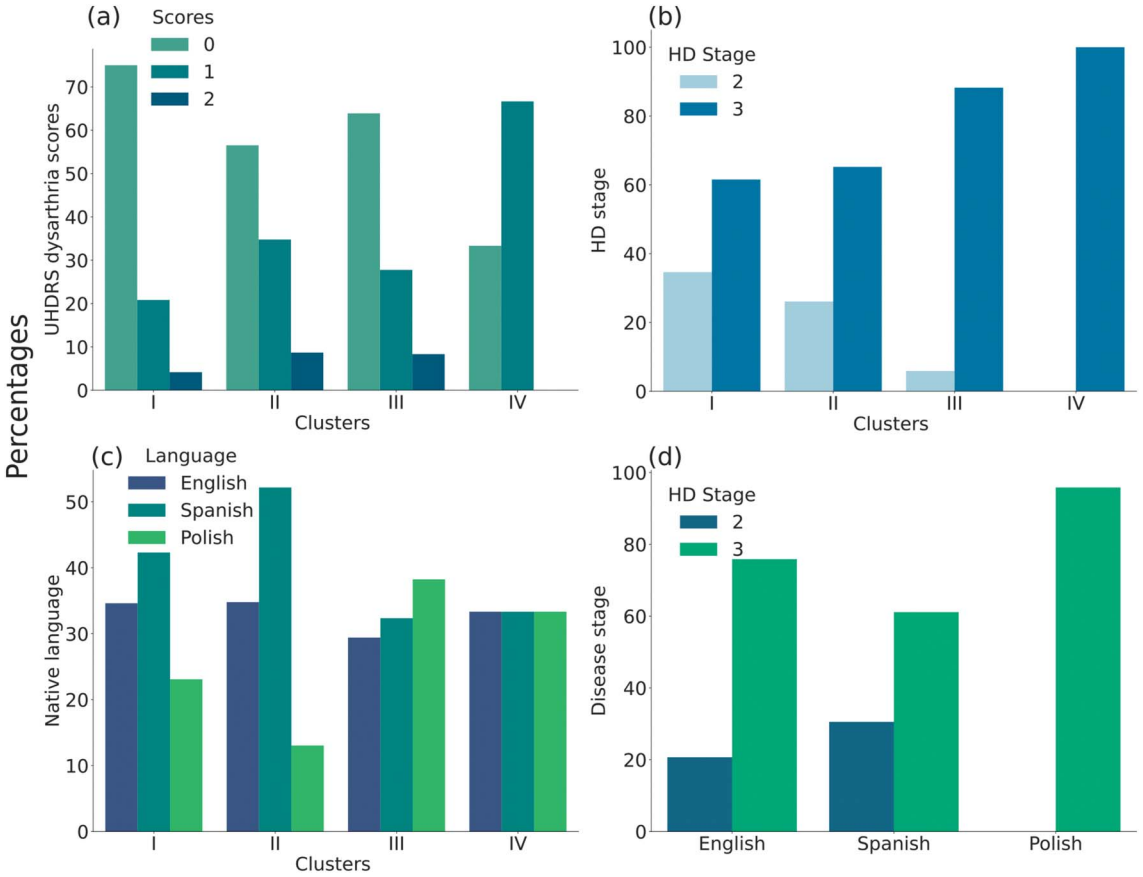


Figure 5. Percentage distribution of (a) Unified Huntington's Disease Rating Scale (UHDRS) dysarthria scores, (b) disease stage, (c) native language within each cluster, and (d) disease stage by language. HD = Huntington's disease.



In addition to capturing specific differences between patient and control groups, while accounting for demographic differences such as age, biological sex, and language, speech biomarkers for neurodegenerative diseases should be generalizable and clinically interpretable. Toward this, the correlation between PCs and acoustic and temporal speech features were investigated. PC1 was associated with harmonicity and noise introduced by incomplete vocal fold closure (HNR), greater variability in the glottal periods (jitter), and expiratory flows (shimmer) and instability of pace (intDur: SR1, SR2, SR3) and rhythm in vocalizations (maxSylRep: SR1, SR2, and SR3; see Table 2). PC2 and PC5 were also related to greater variability in the glottal periods (jitter: SR1, SR2; RAP: SR2; DDP:SR2, SR3), and PC3 was inversely related to speech timing (net speech rate: SR1, SR2, SR3, SR4). PC4 was related with instability of the vocal tract (*SD* MFCC: SR1).

Thus, the participants with HD who compose Cluster I are marked with some variability and instability of the vocal tract (lowest PC1 and mid-to-low range of PC5). Cluster II had similar characteristics but with a greater variability in the vocal tract when compared to Cluster I (higher PC1), lower impairment of speech timing (lower PC3), and the least instability of the vocal tract (lowest PC4). Interestingly, Cluster III comprised the population with lower values for PC2, which corresponds to less instability of the vocal tract, but with moderate pace instability during SR and a higher level of voice aperiodicity, marked by higher number of voice breaks during SV. Finally, Cluster IV has the highest overall indications of loss of harmonicity, greater variability, and instability of the vocal tract as well as greater aperiodicity in the voice.

In addition to a progressive increase in speech impairment from Cluster I to Cluster IV, clinical scores UHDRS total motor score, voluntary movement score, involuntary movement score, and disease burden score also increased progressively from Cluster I to Cluster IV. Consistent with this, total functional capacity and cUHDRS exhibited the reverse trend (see Figure 4).

Approximately, 80% of individuals in Cluster I had a UHDRS dysarthria score of 0. This percentage decreased moving progressively through the clusters as the

proportion of participants with a score of 1 increased. Participants with a score of 2 were present in Clusters I and III, with a higher percentage on Cluster III, though they represent a smaller percentage of the overall cohort (see Figure 5a). When observing disease stage, participants in Stage 2 were more prevalent in Cluster I and decrease through the clusters as the proportion of participants in Stage 3 increased (see Figure 5b). Four distinct clusters were also identified in a previous study of speech features in HD, which used rating scores from expert listeners to characterize the speech patterns (Diehl et al., 2019), in contrast to the automated analysis applied here. The clusters identified in the present study exhibited similar features to those in the study of Diehl et al. (2019), characterized by varied levels of severity, speech rate, and dysarthria, though a group with abnormally fast speech rate was not observed in the present study, which is in line with previous findings in HD (Hartelius et al., 2003; Rusz, Klempíř, et al., 2014; Skodda et al., 2014; Vogel et al., 2012). Together, the results illustrate the ability of the proposed protocol and analysis methods to objectively identify and quantify pathological changes in speech in HD across multiple languages. Four distinct clusters were identified based on the measured speech features, corresponding to different levels of speech impairment and associated with different levels of motor impairment, disease stage, and clinical dysarthria score.

A number of study limitations should be considered when interpreting the results. Speech was recorded using a different device for Spanish speakers than for the English and Polish speakers. Due to device being highly correlated with clinical sites and languages, it could not be included as a covariate in the statistical model, thus not accounting for devices specifically. The position of device on the desk in front of participants could also add a source of error to measurements, in particular to amplitude-based features, including shimmer, as head-mounted microphones are recommended for use with participants with involuntary movements (Patel et al., 2018; Svec & Granqvist, 2010). The recording protocol was chosen to enable the use of mobile devices, typically present within the clinic, and device position was kept consistent across sites to minimize the errors. The inclusion of only the SV and SR tasks in the clustering analysis, with the RP deemed unsuitable for multilanguage approaches, meant that important features of longer stretches of speech and unstructured speech were not captured. Longer stretches of speech, including RP and monologue, can provide a better evaluation of functional capacity, whereas the presented protocol focused on motor function. Future work should include acoustic features that are more robust to multilanguage approaches. The identification of new features that can capture disease-related variations in HD but are robust to differences in languages for longer stretches of speech are needed to reliably include RP and monologue in the analysis. While the features identified have been shown to be insensitive to language during SV and SR tasks for English, Spanish, and Polish speakers, representing Germanic, Romance, and Slavic languages, further extension of the approach is required to establish whether these features are similarly robust across other Indo-European languages and other language families.

Further development of this work can support the application of digital biomarkers as measures of clinical outcomes for patients. Such biomarkers could help characterize dysarthria in HD and enable neurologists and speech and language therapists to conduct automated speech assessments in clinic and longitudinal monitoring of speech in the home. When embedded in a mobile device application, the automated speech assessments could be readily deployed and reproduced with a mobile device to complement standard clinical assessments. Such methods may allow automated quantification of within-patient changes over time or comparison across patients, though the reliability and robustness of the features should be first established. To facilitate the reproducibility of the methods presented, the authors recommend using mobile devices with Android 12 or later, a sampling rate of at least 44.1 kHz, and storage of the recorded speech data in an uncompressed .wav format.

Conclusions

This study demonstrates the potential of acoustic features to automatically characterize speech in HD across different languages, when recorded using mobile devices. Differences were observed between participants with HD and control participants independent of language for a range of acoustic and temporal features during SV and SR tasks. While differences were observed between groups for features from passage reading tasks, these were also influenced by language. Identification of the language-independent features enabled clustering of PCs of the features from the patients with HD into subgroups with different levels of speech and motor impairment. The methods presented can be used with devices already present in the clinic to record speech, automatically estimate acoustic speech features, and categorize patient speech into one of the subgroups defined. The use of acoustic features that have been shown to differ between patients and control participants enables subtle changes in articulation or voice instability to be captured and quantified using automated methods. Each subgroup identified is associated with particular acoustic characteristics and can provide information about the type of speech impairment for that participant. As each subgroup can be related to severity of disease and impairment, the approach could potentially be used to track disease progression or the response to treatment. The proposed method for automatic quantification of speech using mobile devices may also enable remote monitoring of

speech impairment, allowing the speech-language pathologist and clinician to track changes between clinical visits. Used in combination with monitoring of motor symptoms, the approach may provide a powerful biomarker for clinical trials aiming to treat and improve prognosis of HD.

Data Availability Statement

The data sets generated and/or analyzed during the current study are not publicly available in line with research ethics and data protection agreements for this study but are available from the corresponding author on reasonable request.

Acknowledgments

Multi-Domain Lifestyle Targets for Improving Prognosis in Huntington's disease (DOMINO-HD) is funded through the EU Joint Program for Neurodegenerative Disease Research (JPND) as part of the JPND funding call in to Health and Social Care (2019) with funding from Alzheimer's Society, Secretary of State for Health and Social Care, Health and Care Research Wales, Public Health Agency Northern Ireland, Jacques and Gloria Gossweiler Foundation, Bundesministerium für Bildung und Forschung, Narodowe Centrum Badań i Rozwoju, Swiss National Science Foundation (32ND30_185548), and Health Research Board (JPND-HSC-2018-003), awarded to Monica Busse, Madeleine Lowery, Grzegorz Witkowski, and Esther Cubo. The authors would like to acknowledge all the participants who helped and volunteered their time to complete this study. They would also like to acknowledge Lorraine Craig, Kim Munnery, Eileen Donovan, Jennifer Hoblyn, and Muthukumaran Thangaramanujam for assisting with protocol and recruiting; the DOMINO-HD Consortium; Bloomfield Health Services; and the Neuromuscular Systems Group. Biosamples and data used in this work were generously provided by the participants in the Enroll-HD study and made available by CHDI Foundation, Inc. Enroll-HD is a clinical research platform and longitudinal observational study for Huntington's disease families intended to accelerate progress toward therapeutics; it is sponsored by Cure Huntington's Disease Initiative (CHDI) Foundation, a nonprofit biomedical research organization exclusively dedicated to collaboratively developing therapeutics for Huntington's disease. Enroll-HD would not be possible without the vital contribution of the research participants and their families.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). *Fitting linear mixed-effects models using lme4*. ArXiv Preprint. https://doi.org/10.48550/arXiv.1406.5823

Behrman, A. (2021). *Speech and voice science* (4th ed.) Plural.

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences, 17*(1193).

Boersma, P., & Weenink, D. (2022). *Praat: Doing phonetics by computer*. http://www.praat.org

Boyce, S., Fell, H., & MacAuslan, J. (2012). SpeechMark: Landmark detection tool for speech analysis. *13th Annual Conference of the International Speech Communication Association 2012, INTERSPEECH 2012, 3,* 1892–1895. https://doi.org/10.21437/interspeech.2012-513

Brookes, M. (1997). *Voicebox: Speech processing toolbox for MATLAB* [Computers software]. https://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html

Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics, 3*(1), 1–27.

Clinch, S. P., Busse, M., Lelos, M. J., & Rosser, A. E. (2018). Rethinking functional outcome measures: The development of a novel upper limb token transfer test to assess basal ganglia dysfunction. *Frontiers in Neuroscience, 12,* Article 366. https://doi.org/10.3389/fnins.2018.00366

Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-1*(2), 224–227. https://doi.org/10.1109/TPAMI.1979.4766909

Diehl, S. K., Mefferd, A. S., Lin, Y.-C., Sellers, J., McDonell, K. E., de Riesthal, M., & Claassen, D. O. (2019). Motor speech patterns in Huntington disease. *Neurology, 93*(22), E2042–E2052. https://doi.org/10.1212/WNL.0000000000008541

Duffy, J. R. (2005). *Motor speech disorders: Differential diagnosis and management*. Mosby.

Fahed, V. S., Doheny, E. P., Busse, M., Hoblyn, J., & Lowery, M. M. (2022). Comparison of acoustic voice features derived from mobile devices and studio microphone recordings. *Journal of Voice*. Advance online publication. https://doi.org/10.1016/j.jvoice.2022.10.006

Fairbanks, G. (1960). The rainbow passage. *Voice and Articulation Drillbook, 2,* 127.

FDA-NIH Biomarker Working Group. (2016). *FDA-NIH Biomarker Working Group BEST (Biomarkers, EndpointS, and Other Tools) Resource*.

Goedemans, R., & van der Hulst, H. (2013). Rhythm types. *The World Atlas of Language Structures Online*.

Hamilton, A., Ferm, U., Heemskerk, A.-W., Twiston-Davies, R., Matheson, K. Y., Simpson, S. A., & Rae, D. (2012). Management of speech, language and communication difficulties in Huntington's disease. *Neurodegenerative Disease Management, 2*(1), 67–77. https://doi.org/10.2217/nmt.11.78

Hartelius, L., Carlstedt, A., Ytterberg, M., Lillvik, M., & Laakso, K. (2003). Speech disorders in mild and moderate Huntington disease: Results of dysarthria assessments of 19 individuals. *Journal of Medical Speech-Language Pathology, 11*(1), 1–15.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology, 24*(6), 417–441. https://doi.org/10.1037/h0071325

Huntington Study Group. (1996). Unified Huntington's Disease Rating Scale: Reliability and consistency. *Movement Disorders, 11*(2), 136–142. https://doi.org/10.1002/mds.870110204

Kaiser, J. F. (1990). On a simple algorithm to calculate the 'energy' of a signal. *Proceedings of International Conference on Acoustics, Speech and Signal Processing*, *1*(10), 381–384. https://doi.org/10.1109/icassp.1990.115702

Kaploun, L. R., Saxman, J. H., Wasserman, P., & Marder, K. (2011). Acoustic analysis of voice and speech characteristics in presymptomatic gene carriers of Huntington's disease: Biomarkers for preclinical sign onset? *Journal of Medical Speech-Language Pathology, 19*(2), 49–63.

Kim, D., Diehl, S., de Riesthal, M., Tjaden, K., Wilson, S. M., Claassen, D. O., & Mefferd, A. S. (2022). Dysarthria subgroups in talkers with Huntington's disease: Comparison of two data-driven classification approaches. *Brain Sciences, 12*(4), Article 492. https://doi.org/10.3390/brainsci12040492

Kouba, T., Frank, W., Tykalova, T., Mühlbäck, A., Klempíř, J., Lindenberg, K. S., Landwehrmeyer, G. B., & Rusz, J. (2023). Speech biomarkers in Huntington's disease: A cross-sectional study in pre-symptomatic, prodromal and early manifest stages. *European Journal of Neurology, 30*(5), 1262–1271. https://doi.org/10.1111/ene.15726

Kováč, D., Mekyska, J., Aharonson, V., Harár, P., Galaz, Z., Rapcsak, S., Orozco-Arroyave, J. R., Brabenec, L., & Rektorova, I. (2022). *Exploring language-independent digital speech biomarkers of hypokinetic dysarthria*. medRxiv. https://doi.org/10.1101/2022.10.24.22281459

Maddieson, I. (2013a). Syllable structure. *The World Atlas of Language Structures Online*.

Maddieson, I. (2013b). Tone. *The World Atlas of Language Structures Online*.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku P., Ilmoniemi, R. J., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385*(6615), 432–434. https://doi.org/10.1038/385432a0

Orozco-Arroyave, J. R., Hönig, F., Arias-Londoño, J. D., Vargas-Bonilla, J. F., Daqrouq, K., Skodda, S., Rusz, J., & Nöth, E. (2016). Automatic detection of Parkinson's disease in running speech spoken in three different languages. *The Journal of the Acoustical Society of America, 139*(1), 481–500. https://doi.org/10.1121/1.4939739

Patel, R. R., Awan, S.N., Barkmeier-Kraemer, J., Courey, M., Deliyski, D., Eadie, T., Paul, D., Švec, J. G., & Hillman, R. (2018). Recommended protocols for instrumental assessment of voice: American Speech-Language-Hearing Association expert panel to develop a protocol for instrumental assessment of vocal function. *American Journal of Speech-Language Pathology, 27*(3), 887–905. https://doi.org/10.1044/2018_AJSLP-17-0009

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

Penney, J. B., Jr., Vonsattel, J.-P., Macdonald, M. E., Gusella, J. F., & Myers, R. H. (1997). CAG repeat number governs the development rate of pathology in Huntington's disease. *Annals of Neurology, 41*(5), 689–692. https://doi.org/10.1002/ana.410410521

Pettorino, M., Gu, W., Półrola, P., & Fan, P. (2017). Rhythmic characteristics of Parkinsonian speech: A study on Mandarin and Polish. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, August 2017, 3172–3176. https://doi.org/10.21437/Interspeech.2017-850

Ramanarayanan, V., Lammert, A. C., Rowe, H. P., Quatieri, T. F., & Green, J. R. (2022). Speech as a Biomarker: Opportunities, interpretability, and challenges. *Perspectives of the ASHA Special Interest Groups, 7*(1), 276–283. https://doi.org/10.1044/2021_persp-21-00174

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*(C), 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Rusz, J., Klempíř, J., Baborová, E., Tykalová, T., Majerová, V., Čmejla, R., Růžička, E., & Roth, J. (2013). Objective acoustic quantification of phonatory dysfunction in Huntington's disease. *PLOS ONE, 8*(9). https://doi.org/10.1371/journal.pone.0065881

Rusz, J., Klempíř, J., Tykalová, T., Baborová, E., Čmejla, R., Růžička, E., & Roth, J. (2014). Characteristics and occurrence of speech impairment in Huntington's disease: Possible influence of antipsychotic medication. *Journal of Neural Transmission, 121*(12), 1529–1539. https://doi.org/10.1007/s00702-014-1229-8

Rusz, J., Saft, C., Schlegel, U., Hoffman, R., & Skodda, S. (2014). Phonatory dysfunction as a preclinical symptom of Huntington disease. *PLOS ONE, 9*(11), 1–7. https://doi.org/10.1371/journal.pone.0113412

Rusz, J., Tykalova, T., Ramig, L. O., & Tripoliti, E. (2021). Guidelines for speech recording and acoustic analyses in dysarthrias of movement disorders. *Movement Disorders, 36*(4), 803–814. https://doi.org/10.1002/mds.28465

Satopää, V., Albrecht, J., Irwin, D., & Raghavan, B. (2011). Finding a "kneedle" in a haystack: Detecting knee points in system behavior. *Proceedings of the International Conference on Distributed Computing Systems*, 166–171. https://doi.org/10.1109/ICDCSW.2011.20

Schobel, S. A., Palermo, G., Auinger, P., Long, J. D., Ma, S., Khwaja, O. S., Trundell, D., Cudkowicz, M., Hersch, S., Sampaio, C., Dorsey, E. R., Leavitt, B. R., Kieburtz, K. D., Sevigny, J. J., Langbehn, D. R., Tabrizi, S. J., TRACK-HD, COHORT, CARE-HD, & 2CARE Huntington Study Group Investigators. (2017). Motor, cognitive, and functional declines contribute to a single progressive factor in early HD. *Neurology, 89*(24), 2495–2502. https://doi.org/10.1212/WNL.0000000000004743

Skodda, S., Flasskamp, A., & Schlegel, U. (2010). Instability of syllable repetition as a model for impaired motor processing: Is Parkinson's disease a "rhythm disorder"? *Journal of Neural Transmission, 117*(5), 605–612. https://doi.org/10.1007/s00702-010-0390-y

Skodda, S., Schlegel, U., Hoffmann, R., & Saft, C. (2014). Impaired motor speech performance in Huntington's disease. *Journal of Neural Transmission, 121*(4), 399–407. https://doi.org/10.1007/s00702-013-1115-9

Styler, W. (2017). On the acoustical features of vowel nasality in English and French. *The Journal of the Acoustical Society of America, 142*(4), 2469–2482. https://doi.org/10.1121/1.5008854

Svec, J. G., & Granqvist, S. (2010). Guidelines for selecting microphones for human voice production research. *American Journal of Speech-Language Pathology, 19*(4), 356–368. https://doi.org/10.1044/1058-0360(2010/09-0091)

Tsanas, A., & Arora, S. (2021). Assessing Parkinson's disease speech signal generalization of clustering results across three countries: Findings in the Parkinson's voice initiative study. *Proceedings of the 14th International Conference on Bio-Inspired Systems and Signal Processing; Part of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2021*, *4*(Biostec), 124–134. https://www.scitepress.org/Papers/2021/103830/103830.pdf [PDF]

Tsanas, A., Little, M. A., McSharry, P. E., & Ramig, L. O. (2011). Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity. *Journal of the*

Royal Society Interface, 8(59), 842–855. https://doi.org/10.1098/rsif.2010.0456

Vogel, A. P., Churchyard, A., & Stout, J. (2016). Two different phenomena in basic motor speech performance in premanifest Huntington disease. *Neurology, 87*(21), 2283. https://doi.org/10.1212/WNL.0000000000003392

Vogel, A. P., Shirbin, C., Churchyard, A. J., & Stout, J. C. (2012). Speech acoustic markers of early stage and prodromal Huntington's disease: A marker of disease onset? *Neuropsychologia, 50*(14), 3273–3278. https://doi.org/10.1016/j.neuropsychologia.2012.09.011

Winter, B. (2013). *Linear models and linear mixed effects models in R with linguistic applications*. ArXiv Preprint. https://doi.org/10.48550/arXiv.1308.5499

Yorkston, K. M., Beukelman, D. R., & Tice, R. (1996). *Sentence Intelligibility Test (SIT)*. Tice Technology Services.