Articulatory Coordination for Speech Motor Tracking in Huntington Disease

Matthew Perez[1], Amrit Romana[1], Angela Roberts[2], Noelle Carlozzi[3], Jennifer Ann Miner[3], Praveen Dayalu[4], Emily Mower Provost[1]

[1]Computer Science and Engineering, University of Michigan, Ann Arbor, Michigan, USA
[2]Communication Sciences and Disorders, Northwestern University, Evanston, Illinois, USA
[3]Physical Medicine & Rehabilitation, University of Michigan, Ann Arbor, MI
[4]Michigan Medicine, University of Michigan, Ann Arbor, MI

mkperez@umich.edu, aromana@umich.edu, angela.roberts@northwestern.edu,
carlozzi@med.umich.edu, jenminer@med.umich.edu, pravd@med.umich.edu, emilykmp@umich.edu

Abstract

Huntington Disease (HD) is a progressive disorder which often manifests in motor impairment. Motor severity (captured via motor score) is a key component in assessing overall HD severity. However, motor score evaluation involves in-clinic visits with a trained medical professional, which are expensive and not always accessible. Speech analysis provides an attractive avenue for tracking HD severity because speech is easy to collect remotely and provides insight into motor changes. HD speech is typically characterized as having irregular articulation. With this in mind, acoustic features that can capture vocal tract movement and articulatory coordination are particularly promising for characterizing motor symptom progression in HD. In this paper, we present an experiment that uses Vocal Tract Coordination (VTC) features extracted from read speech to estimate a motor score. When using an elastic-net regression model, we find that VTC features significantly outperform other acoustic features across varied-length audio segments, which highlights the effectiveness of these features for both short- and long-form reading tasks. Lastly, we analyze the F-value scores of VTC features to visualize which channels are most related to motor score. This work enables future research efforts to consider VTC features for acoustic analyses which target HD motor symptomatology tracking.

Index Terms: Huntington disease, motor impairment, vocal tract coordination, articulatory coordination, acoustic features, motor symptom tracking

1. Introduction

Huntington Disease (HD) is a genetic, neurodegenerative disease that affects approximately 1 out of 10,000 individuals. Those afflicted with HD develop motor, cognitive, and psychiatric problems which worsen over time. Although there is no known cure for HD, having the ability to track symptomatology is imperative in medical efforts for developing effective treatments. Quantitative speech analysis provides a promising avenue for assisting medical professionals in characterizing symptom progression, since motor speech deficits are one of the most common symptoms observed in HD patients, occurring in roughly 90% of cases [1, 2, 3, 4, 5, 6].

Manifest HD can cause a variety of motor-related symptoms in speech resulting in reduced speech rate, abnormal speech rhythm, increased pauses, shorter vowel durations, higher intervocalization duration, and reduced articulation [2, 3, 4, 5]. These effects can also be compounded by co-morbidities, such as dysarthria (slurred speech) and/or bradykinesia (slow movement), which can impact speech production depending on the elicitation task. In premanifest HD, speech trends such as increased articulation rate, imprecise pace performance, and higher variability in syllable repetition are observed [7]. These clinical findings in manifest and premanifest HD suggest that speech can be a useful signal for disease progression tracking. With this in mind, future applications should focus on automatic HD analyses that capture a wide range of speech motor irregularities over HD severity.

Prior works have focused on differentiating between healthy control speakers and those with premanifest and manifest HD [8, 9, 10]. However, as we consider applications that can help medical professionals monitor symptom progression, it is important to consider severity on a continuous scale, such as predicting a motor score. There are several continuous scores useful for severity tracking; however, in this work we specifically focus on Total Motor Score (TMS) because it represents a critical component of the standard assessment used by medical professionals to evaluate HD severity [11].

Acoustic features that capture irregular articulatory movement are ideal for HD motor tracking based on the clinical literature regarding HD speech [2]. Vocal Tract Coordination (VTC) features are strong candidates for this task by quantifying articulatory coordination using correlations across low-level acoustic features. Although previous research has used these features to capture psychomotor articulatory coordination in depression [12, 13], these features have not been explored at capturing the broad range of motor symptoms present in HD speech.

In this work, we investigate and analyze the effectiveness of VTC features for HD symptom progression tracking. We experiment with two extraction heuristics for VTC, eigendecomposition- (EVTC) and full-VTC (FVTC), in order to study the effect of dimensionality reduction on these features when characterizing motor symptomatology. We ultimately show that FVTC features achieve performance improvements over OpenSMILE, raw acoustic features, and EVTC when used for predicting TMS on read speech (i.e., the Grandfather Passage). We also investigate how the length of the audio sample is related to downstream performance and analyze which FVTC feature channels are most relevant to motor score tracking.

2. Related Work

Prior work by Riad et. al. have shown that phonatory features are useful for classifying between control, premanifest, and manifest HD when using acoustic data from a max phonation task [10]. These phonatory features include jitter/shimmer,

probability of voicing, and statistics over mel-cepstral features. The authors also show that these phonatory features achieve an R^2 of 0.53 when predicting TMS. However, questions still remain as to whether these phonatory features can be extended to tasks that involve complex speech production (i.e., passage readings or spontaneous speech), which is the ideal for remote health monitoring applications. In our work, we extract similar phonatory and rhythm features using the OpenSMILE toolkit.

Previous work by Perez et. al. looked at capturing pauses, speech rate, and pronunciation from generated transcripts to automatically differentiate between healthy controls and gene positive HD participants who read the Grandfather Passage [8]. Their results show that neural networks could classify HD with up to 87% accuracy and that the effectiveness of pauses, speech rate, and pronunciation were correlated with disease severity. However, a limitation of this approach is that the extracted features are directly tied to the performance of ASR systems, which typically underperform in unrestricted, disordered speech settings [14, 15, 16]. In our current work, we remove this dependence on ASR systems for feature extraction by extracting acoustic features directly from audio.

Recent work by Romana et. al. looked at using vowel distortion to classify manifest vs. premanifest HD [9]. The authors used manually extracted vowel segments from the Grandfather Passage. The authors demonstrate that these features are not only correlated to TMS but achieve 80% classification accuracy of premanifest vs. manifest HD using trained models. Classification was performed at the speaker-level so features were computed over the reading passage. In contrast, rather than restricting to manually extracted vowels, the current work investigates acoustic features that can be extracted automatically over the entire passage.

3. Data

The dataset we use was collected at the University of Michigan [17]. The dataset consists of 62 English-speaking participants (31 healthy, 31 with HD). Healthy control participants had no history of either neurological disorders or speech impairments. Individuals with HD had to have a positive gene test and/or clinical HD diagnosis by a physician.

The motor severity of each participant was assessed using the TMS, the motor subsection of the Unified Huntington Disease Rating Scale (UHDRS) [11], a standard assessment used by medical professionals to evaluate HD severity. The TMS is a holistic motor evaluation, which sums together individual motor subsections that target eye, trunk, gait, tongue, and speech movement. The TMS ranges from 0 (healthy motor functionality) to 128 (severe motor impairment). In our dataset, our HD speakers have an average TMS of 33.6 ($\pm$ 23.6), while our Control speakers have an average TMS of 4.4 ($\pm$ 2.7).

HD participants were designated with premanifest HD if they had a diagnostic confidence rating of <4 on the last item of the TMS, while manifest HD was defined by a confidence rating of 4 on the last item of the TMS [18]. Within the manifest group, we label participants as early- or late-stage based on their Total Functional Capacity (TFC) scores, which range from 0 (low functioning) to 13 (high functioning) [19]. Manifest early-stage was defined by a TFC score of 7-13, while manifest late-stage was defined by a TFC score of 0-6 [20]. The speaker breakdown according to HD severity is: 12 premanifest HD, 12 manifest early-stage, and 7 manifest late-stage.

The data include both interviews and a series of tasks [17]. All tasks were recorded using a table microphone at 44.1 kHz.

This study focuses on the reading portion, where participants read the Grandfather Passage [21]. This reading task contains 129 words and 169 syllables and is commonly used to test for dysarthria in speech-language pathology.

4. Features

4.1. Vocal Tract Coordination Features

Although previous works have shown success at using VTC-derived features to capture psychomotor articulatory differences when characterizing depression [13, 22], little work has been done on extending similar correlation-based features towards other domains. Clinical research has not only characterized HD speech as having irregular articulation but has even suggested that motor performance of the vocal tract and trunk extremities is controlled by the same mechanisms, presumably the basal ganglia [2, 6]. This highlights the vocal tract as a key area of study for motor symptom tracking in HD and motivates the investigation of VTC features for characterizing the wide- range of articulatory symptoms in HD.

VTC features are derived from acoustic signals. In our work we use librosa to extract 16-channel mel-frequency cepstral coefficients (mfcc), delta mel-frequency cepstral coefficients (dmfcc) using a hamming window of 25 ms with a 10 ms step size. We remove the 0th coefficient and apply speaker-level cepstral mean and variance normalization.

We calculate VTC features by applying auto- and cross-correlation functions to time-delayed channels (i.e., feature dimension) of a given acoustic signal, following the same approach as outlined in [13]. For a given utterance, a normalized correlation $r_{i,j}^d$ is computed over the acoustic signal x following equation 1, where i and j are channels and d is the time-delay used to shift a given channel.

$$r_{i,j}^d = \frac{\sum_{t=1}^{T-d} x_i[t]x_j[t+d]}{\sqrt{r_{i,i}^0 r_{j,j}^0}} \qquad (1)$$

We experiment with two feature formats for VTC. The first uses the full feature vector, which we call Full-VTC (FVTC). For FVTC, we concatenate all $r_{i,j}^d$, which results in a $N \times N \times D$ feature vector, where N is the number of channels (15) across the entire utterance and a max delay of D (80). The second format applies eigendecomposition across the channels of FVTC, which we call eigen-VTC (EVTC). EVTC was originally used in previous works [22, 12] and results in a $N \times D$ feature vector. Our feature extraction code will be made publicly available [1].

4.2. Baseline Features

Prior research has shown promise in using IS10 features from the OpenSMILE toolkit for emotion recognition, cognitive decline assessment, and characterizing disordered speech [23, 24, 25]. We extract IS10 features which include low-level mfcc with appended deltas, shimmer, jitter, log-mel frequency band, and fundamental frequency features. Another benefit of IS10 is that it also contains some features which were used in [10]. For each utterance, we compute broad statistics (mean, std, min, max, and median) over the frame-level feature matrix leading to a 352-dim feature vector.

For our second set of baseline features (raw acoustic features), we extract and use low-level acoustic features (mfcc and dmfcc) and compute broad statistics over the utterance in order

---

[1]https://github.com/matthewkperez/VTC-features

Table 1: *Elastic-net regression task when predicting Total Motor Score at the speaker-level using 10s segments. Results are averaged over 100 runs ($\pm$ std). † indicates that the marked performance is significantly higher than best baseline feature (raw-dmfcc). Significance is assessed at $p < 0.05$ using the Tukey's honest test on the ANOVA statistics.*

| Features	RMSE	R^2	CCC	CCC-pre	CCC-early	CCC-late
Transcript Feats	18.0 (4.4)	0.32 (0.29)	0.52 (0.19)	0.28	0.22	0.09
OpenSMILE IS10	20.3 (3.7)	0.16 (0.34)	0.39 (0.21)	0.25	0.2	0.06
Raw acoustic features						
mfcc	20.5 (4.0)	0.16 (0.29)	0.39 (0.19)	0.08	0.12	0.07
dmfcc	19.9 (3.9)	0.21 (0.29)	0.43 (0.17)	0.21	0.04	0.17
EVTC						
mfcc	20.5 (4.1)	0.17 (0.26)	0.38 (0.18)	0.18	0.09	0.19
dmfcc	20.5 (4.0)	0.13 (0.4)	0.38 (0.21)	0.24	0.21	0.24
FVTC						
mfcc	22.1 (3.1)	0.04 (0.29)	0.29 (0.19)	0.2	-0.2	0.12
dmfcc	**17.9 (3.5)** †	**0.32 (0.36)**	**0.51 (0.2)**†	**0.26**	**0.26**	**0.35**
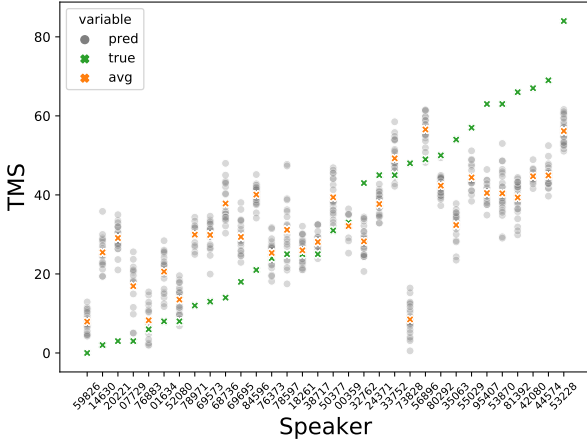
Figure 1: *Elastic-net model using FVTC-dmfcc features, where the top-75 features. 10s audio segments were used and the model was run a 100 times.*

to generate a fixed feature vector. We compute mean, std, min, max, and median following that of previous works [23, 26].

5. Experiment

For all experiments, we initially segment the audio data into smaller chunks using a sliding window of 10s in order to increase the number of training samples in our dataset (Section 5.1) and then explore potential performance changes as a function of segment length (Section 5.2). We assign the same TMS score to all segments from a given speaker. We evaluate performance using a randomly selected 20% held-out, speaker-independent test set. To account for randomness in test set selection, model initialization, and training, we run each experiment 100 times and present metric averages (± 1 std) over all runs. We downsample the control population to a randomly selected seven speakers due to the large TMS imbalance (i.e. majority of speakers are control or premanifest). This selection was motivated by the number of late-stage HD speakers (seven) and serves to ultimately circumvent biasing the classifier towards low TMS scores.

We perform feature normalization and select the top 75-features (the upper bound for our raw acoustic feature set) based on F-value scores using the training set. Selected features are then used to train an elastic-net model, where L1 and L2 regularization is combined using default hyperparameters, (C=1 and ratio=0.5) [27]. Elastic-net is utilized for our TMS regression task due to its ability to handle multicollinearity and perform variable selection and regularization simultaneously [10, 26].

5.1. TMS Score Prediction

We use the elastic-net model to predict TMS for each segment and perform speaker-level averaging over all segments for the final TMS prediction. We evaluate system performance using root-mean-square error (RMSE), R^2, and concordance correlation coefficient (CCC) between the predicted and target scores. In addition, we compute the CCC for each HD severity (pre-manifest, early, and late) over all runs. With this, we can better understand what types of speakers the model is able to fit with regards to HD severity. To assess the statistical significance we perform a Tukey honest significance test using $p < 0.05$.

Our goal is to identify how well VTC features (FVTC and EVTC) characterize TMS in comparison to other baseline feature sets. Our results presented in Table 1 shows that FVTC-dmfcc outperforms all baseline methods across all evaluation metrics when estimating TMS. We note significant improvements in RMSE and CCC, achieving average values of 17.9 and 0.51 respectively. We also note that when comparing EVTC and FVTC, eigendecomposition seems to benefit mfcc; however, for dmfcc using FVTC features provides better performance. Ultimately, these results highlight FVTC-dmfcc as a promising acoustic biomarker for modeling motor symptomatology.

Figure 1 illustrates the performance of FVTC-dmfcc for all speakers by showing the ground truth (green), speaker prediction (grey), and mean prediction (orange) TMS. One point of emphasis is the model's performance for higher TMS, where we can see a stronger correlation (CCC = 0.35) for late-stage speakers, compared to premanifest- and early-stage speakers (Table 1). One outlier to note is speaker 73828, which the model incorrectly estimates as having low TMS. Although further investigation is required, initial analyses show that this speaker is an outlier across all feature sets, which seems to suggest that this speaker's motor issues may not manifest in speech.
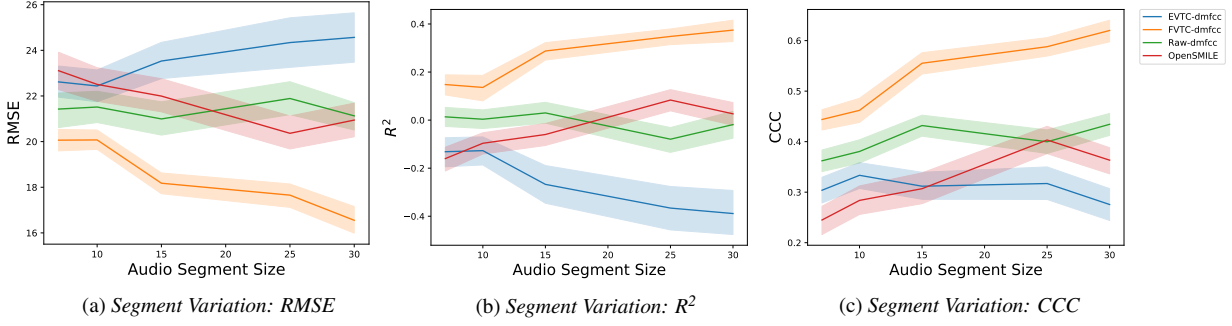
(a) *Segment Variation: RMSE* (b) *Segment Variation: R^2* (c) *Segment Variation: CCC*

Figure 2: *RMSE, R^2, and CCC using varying segment lengths to splice the Grandfather Passage. TMS estimation is performed at the segment-level.*
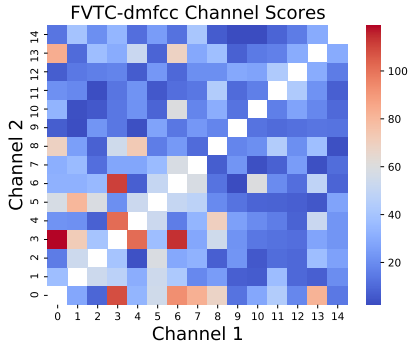

Figure 3: *F-value scores for FVTC-dmfcc using 10s segments averaged over all 100 runs and channel maximum taken over all delays.*

5.2. Segment Size Variation

Our next experiment analyzes how the segment-level performance of our system changes as the length of the segments is varied. In the previous analysis, we restricted segments to 10s due to the relatively small dataset size. However, in this section, we analyze potential performance increases resulting from longer segment lengths to inform future studies. We segment the grandfather passage into 7, 10, 15, 20, 25, and 30 second segments and perform TMS estimation at the segment-level.

Figure 2 shows the performance of RMSE, R^2, and CCC for our different feature sets when varying audio segment size. We can see performance improvement in RMSE, R^2, and CCC when using FVTC-dmfcc across all segment sizes. OpenS-MILE, raw feats, and FVTC all show improvements when using longer audio segments. We hypothesize that for OpenSMILE and raw feats this is likely due to taking broad statistics over longer audio sequences. In the case of FVTC, we demonstrate that longer audio intervals allow for more robust and generalizable correlation computation, which is a finding that is also consistent with previous work [13].

These results motivate the need for longer form passage readings since nearly all acoustic features seem to benefit from longer audio. However, in the case of both short- and long-form audio readings, FVTC-dmfcc should be an acoustic feature considered for characterizing TMS.

5.3. FVTC Channel Analysis

Our last experiment analyzes the FVTC-dmfcc feature space using the scores derived from F-value between feature and TMS.

Figure 3 shows a heatmap where the x- and y-axis represent the i and j channels used for the FVTC-dmfcc features. Feature scores are averaged over all 100 iterations and the maximum score is taken for each channel across all delays. Channels 0-6 seems to represent a dense amount of relevant information for TMS regression, accounting for 66% of the final selected features, which suggests that correlations across channels in the low frequency space are related to TMS.

6. Conclusion

In this work we present a novel investigation into the effectiveness of Vocal Tract Coordination (VTC) features for characterizing motor symptomatology in HD speech. We show that FVTC-dmfcc significantly outperforms all other acoustic features in predicting TMS with regards to RMSE and CCC by achieving average values of 17.9, 0.32, and 0.51 respectively. Furthermore, we demonstrate that FVTC-dmfcc outperforms other acoustic features when classifying varied length segments, which suggests these features should be considered when analyzing both short- and long-form reading tasks. We also show that FVTC-dmfcc performance improves as audio length increases, which motivates data collection methods to push for long form audio readings. Lastly, we analyze channel importance by plotting F-value scores and demonstrate that low-frequency channels are most relevant to TMS.

Future work, will encompass further analysis into VTC features and their ability to characterize disordered speech. We plan to investigate other disorders, such as Parkinson's Disease. Additionally, as we start to consider passive, remote health monitoring applications we would like to relax the dependence on read speech tasks and utilize free speech as audio input.

7. Acknowledgements

8. References

[1] A. B. Young, I. Shoulson, J. B. Penney, S. Starosta-Rubinstein, F. Gomez, H. Travers, M. A. Ramos-Arroyo, S. R. Snodgrass, E. Bonilla, H. Moreno *et al.*, "Huntington's disease in venezuela: neurologic features and functional decline," *Neurology*, vol. 36, no. 2, pp. 244–244, 1986.

[2] S. Skodda, U. Schlegel, R. Hoffmann, and C. Saft, "Impaired motor speech performance in huntington's disease," *Journal of Neural Transmission*, vol. 121, no. 4, pp. 399–407, 2014.

[3] J. Rusz, J. Hlavnička, R. Čmejla, and E. Růžička, "Automatic evaluation of speech rhythm instability and acceleration in dysarthrias associated with basal ganglia dysfunction," *Frontiers in bioengineering and biotechnology*, vol. 3, p. 104, 2015.

[4] J. Rusz, J. Klempíř, E. Baborová, T. Tykalová, V. Majerová, R. Čmejla, E. Růžička, and J. Roth, "Objective acoustic quantification of phonatory dysfunction in huntington's disease," *PloS one*, vol. 8, no. 6, p. e65881, 2013.

[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] S. K. Diehl, A. S. Mefferd, 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.

[7] S. Skodda, W. Grönheit, C. Lukas, B. Bellenberg, S. M. von Hein, R. Hoffmann, and C. Saft, "Two different phenomena in basic motor speech performance in premanifest huntington disease," *Neurology*, vol. 86, no. 14, pp. 1329–1335, 2016.

[8] 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 *Interspeech*, vol. 2018. NIH Public Access, 2018, p. 1898.

[9] A. Romana, J. Bandon, N. Carlozzi, A. Roberts, and E. M. Provost, "Classification of manifest huntington disease using vowel distortion measures," *Proc. Interspeech 2020*, pp. 4966–4970, 2020.

[10] R. Riad, H. Titeux, L. Lemoine, J. Montillot, J. H. Bagnou, X.-N. Cao, E. Dupoux, and A.-C. Bachoud-Lévi, "Vocal markers from sustained phonation in huntington's disease," *Proc. Interspeech 2020*, pp. 1893–1897, 2020.

[11] 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.

[12] J. R. Williamson, T. F. Quatieri, B. S. Helfer, G. Ciccarelli, and D. D. Mehta, "Vocal and facial biomarkers of depression based on motor incoordination and timing," in *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge*, 2014, pp. 65–72.

[13] Z. Huang, J. Epps, and D. Joachim, "Exploiting vocal tract coordination using dilated cnns for depression detection in naturalistic environments," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 6549–6553.

[14] H. Christensen, S. Cunningham, C. Fox, P. Green, and T. Hain, "A comparative study of adaptive, automatic recognition of disordered speech," in *Thirteenth Annual Conference of the International Speech Communication Association*, 2012.

[15] M. Perez, Z. Aldeneh, and E. M. Provost, "Aphasic speech recognition using a mixture of speech intelligibility experts," *Proc. Interspeech 2020*, pp. 4986–4990, 2020.

[16] K. C. Fraser, F. Rudzicz, N. Graham, and E. Rochon, "Automatic speech recognition in the diagnosis of primary progressive aphasia," in *Proceedings of the fourth workshop on speech and language processing for assistive technologies*, 2013, pp. 47–54.

[17] N. E. Carlozzi, N. R. Boileau, A. Roberts, P. Dayalu, D. L. Hanifan, J. A. Miner, D. Claassen, and E. M. Provost, "Understanding speech and swallowing difficulties in individuals with huntington disease: Validation of the hdqlife speech difficulties and swallowing difficulties item banks," *Quality of Life Research*, vol. 30, no. 1, pp. 251–265, 2021.

[18] D. Liu, J. D. Long, Y. Zhang, L. A. Raymond, K. Marder, A. Rosser, E. A. McCusker, J. A. Mills, and J. S. Paulsen, "Motor onset and diagnosis in huntington disease using the diagnostic confidence level," *Journal of neurology*, vol. 262, no. 12, pp. 2691–2698, 2015.

[19] I. Shoulson, R. Kurlan, A. J. Rubin, D. Goldblatt, J. Behr, C. Miller, J. Kennedy, K. A. Bamford, E. D. Caine, D. K. Kido *et al.*, "Assessment of functional capacity in neurodegenerative movement disorders: Huntington's disease as a prototype," *Quantification of neurologic deficit. Boston: Butterworths*, pp. 271–283, 1989.

[20] K. Marder, H. Zhao, R. Myers, M. Cudkowicz, E. Kayson, K. Kieburtz, C. Orme, J. Paulsen, J. Penney, E. Siemers *et al.*, "Rate of functional decline in huntington's disease," *Neurology*, vol. 54, no. 2, pp. 452–452, 2000.

[21] J. Reilly and J. L. Fisher, "Sherlock holmes and the strange case of the missing attribution: A historical note on "the grandfather passage"," 2012.

[22] J. R. Williamson, T. F. Quatieri, B. S. Helfer, R. Horwitz, B. Yu, and D. D. Mehta, "Vocal biomarkers of depression based on motor incoordination," in *Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge*, 2013, pp. 41–48.

[23] Y. Pan, B. Mirheidari, Z. Tu, R. O'Malley, T. Walker, A. Venneri, M. Reuber, D. Blackburn, and H. Christensen, "Acoustic feature extraction with interpretable deep neural network for neurodegenerative related disorder classification," *Proc. Interspeech 2020*, pp. 4806–4810, 2020.

[24] Y. Pan, B. Mirheidari, M. Reuber, A. Venneri, D. Blackburn, and H. Christensen, "Improving detection of alzheimer's disease using automatic speech recognition to identify high-quality segments for more robust feature extraction," *Proc. Interspeech 2020*, pp. 4961–4965, 2020.

[25] G. An, D. G. Brizan, M. Ma, M. Morales, A. R. Syed, and A. Rosenberg, "Automatic recognition of unified parkinson's disease rating from speech with acoustic, i-vector and phonotactic features," in *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.

[26] T. Alhanai, R. Au, and J. Glass, "Spoken language biomarkers for detecting cognitive impairment," in *2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*. IEEE, 2017, pp. 409–416.

[27] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.