INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

Cross-Database Models for the Classification of Dysarthria Presence

Stephanie Gillespie[1], Yash-Yee Logan[1], Elliot Moore[1], Jacqueline Laures-Gore[2], Scott Russell[3], Rupal Patel[4]

[1]Georgia Institute of Technology, United States of America
[2]Georgia State University, United States of America
[3]Grady Memorial Hospital, United States of America
[4]Northeastern University, United States of America
{sgillespie6,ylogan3,em80}@gatech.edu, jlaures@gsu.edu, smrussell@gmh.edu, r.patel@neu.edu

Abstract

Dysarthria is a motor speech disorder that impacts verbal articulation and co-ordination, resulting in slow, slurred and imprecise speech. Automated classification of dysarthria subtypes and severities could provide a useful clinical tool in assessing the onset and progress in treatment. This study represents a pilot project to train models to detect the presence of dysarthria in continuous speech. Subsets of the Universal Access Research Dataset (UA-Speech) and the Atlanta Motor Speech Disorders Corpus (AMSDC) database were utilized in a cross-database training strategy (training on UA-Speech / testing on AMSDC) to distinguish speech with and without dysarthria. In addition to traditional spectral and prosodic features, the current study also includes features based on the Teager Energy Operator (TEO) and the glottal waveform. Baseline results on the UA-Speech dataset maximize word- and participant-level accuracies at 75.3% and 92.9% using prosodic features. However, the cross-training of UA-Speech tested on the AMSDC maximize word- and participant-level accuracies at 71.3% and 90% based on a TEO feature. The results of this pilot study reinforce consideration of dysarthria subtypes in cross-dataset training as well as highlight additional features that may be sensitive to the presence of dysarthria in continuous speech.

Index Terms: Speech disorders, dysarthria, cross-database training, glottal features, Teager

1. Introduction

Dysarthria is a neuro-motor impairment that inhibits an individual's verbal articulation and co-ordination, resulting in discordant speech production. Dysarthria can include any disturbance of the basic components underlying speech production, including respiration, phonation, articulations, resonance, and prosody. In spastic dysarthria, for instance, speech is affected by atypical speech rates, incomplete consonant closure and monotonicity [1]. Dysarthria is often a consequence of a neurological trauma (i.e., cerebral palsy, stroke) or a symptom of degenerative disorders (i.e., hypokinetic dysarthria for Parkinson's disease).

Several research efforts have focused on designing automatic speech recognition (ASR) systems to help persons with dysarthria communicate [2-4]. However, the task of automatically diagnosing and classifying the subtype and severity of dysarthria is less common. The traditional approach involves clinicians performing perceptual judgements of speech to appraise the type and severity of the disorder, such as with the Frenchay Dysarthria Assessment [5]. Clinicians perform their analysis via subjective listening tests which are costly, laborious, and prone to internal biases of the examiner [6]. The subjective nature of these tests has raised doubts about the reliability and validity of perceptual judgements by clinicians to consistently differentiate individual and coexistent speech disorders [7]. This motivates the design of a clinical tool that can objectively differentiate dysarthric from healthy speech and distinguish the subtype and severity of the dysarthria to assist with treatments.

Most speech processing work has analyzed dysarthria to broaden the applicability of automatic speech recognition systems. Shahamiri provides a comparison of several recent ASR models for dysarthric speech [4]. The Universal Access Research Dataset (UA-Speech) was collected with the goal of expanding ASR systems to a more diverse population including those with dysarthria [8], and has been used in multiple efforts for ASR [2, 4]. Other English databases of speakers with dysarthria include: Nemours, Torgo, Whitaker and HomeService [9-12]. Work presented by Sriranjani et al. combined datasets without dysarthric speech with the UA-Speech and Nemours Dataset for larger datasets to train ASR systems, concluding incorporation of non-dysarthric data for models of dysarthric speech reduced performance [13]. While the goal of ASR with dysarthric speech has not focused on diagnosing or detecting dysarthria, Laaridh et al. highlight that one approach to automatic detection of intelligibility of dysarthric speech could be the word transcript error rate based on automatic speech transcriptions systems [14].

The goal of this research is to work towards a clinical tool for detecting the presence of dysarthria from vocal acoustics in continuous speech. This could be used as a diagnostic tool on its own, or as a pre-processing tool for research studying other language disorders such as aphasia [15, 16] where up to 50% of adults with aphasia are also diagnosed as having dysarthria [17]. However, dysarthria falls into many subcategories and severity levels. Dysarthria speech databases tend to have few participants and/or examples only representing a small number of subtypes. This work represents the preliminary results of a pilot study to examine the impact of cross-database training to detect dysarthria. To the authors' knowledge, there is little prior work with multiple corpora to automatically classify dysarthric from healthy speech or identify dysarthria subtype and severity; one exception is work by Orozco-Arroyave et al., which focuses on detecting hypokinetic dysarthria for Parkinson's disease across different languages [18].

http://dx.doi.org/10.21437/Interspeech.2017-216

Table 1: *UA-Speech Dataset Demographic and Clinical Information for Dysarthria Participants*

	Females	Males
# of Participants	4	11
Age Range	18-51	18-58
# with Spastic Dysarthria	3	8
# with Flaccid Dysarthria	1	3
# very low, low, mid, high intelligibility	1,1,1,1	3,2,2,4

Table 2: *AMSDC Demographic and Clinical Information for Included Participants*

	Females	Males
# of Participants	22	35
Age Range	38-76	30-79
# with Spastic Dysarthria	4	3
# with Flaccid Dysarthria	4	9
# with Mixed Dysarthria	11	11
# with Other Dysarthria	3	12
# very low, low, mid, high intelligibility	1,2,2,17	1,2,7,25

2. Databases

The UA-Speech dataset [8] is one of the largest, freely-available datasets of dysarthric speech and is used to establish a comparison of the presented work to prior work. The inclusion of speech from healthy controls allows a model to be built to detect the presence of dysarthria. However, the UA-Speech dataset is not sufficiently diverse to create a clinical dysarthria detection tool: the majority of speakers have spastic dysarthria originating from cerebral palsy. To expand the applicability of the proposed tool, the models trained on UA-Speech are tested on the Atlanta Motor Speech Disorder Corpus (AMSDC) [19], containing more dysarthria subtypes and etiology. Without speech from healthy controls, the AMSDC could not be used to train the detection model.

2.1. UA-Speech Dataset

Originally collected by the University of Illinois, the UA-Speech database contains speech recordings from 15 speakers (4 female, 11 male) of mostly spastic dysarthria due to cerebral palsy [8]. Additionally, 13 controls without dysarthria (4 female, 9 male) were included for a total of 28 participants. Subjects were asked to read single isolated words shown on a laptop screen, with words representing the 10 digits, 26 radio alphabet letters, computer commands, common words from the Brown corpus of written English, and uncommon words from children's novels selected to maximize phone-sequence diversity, for a total of 765 isolated words, 455 of which were distinct. An 8-microphone array was used to collect the speech at 48 kHz, and processed to create 7 channels. In this work, one channel was randomly selected per participant to be analyzed to reduce repetitions of an individual's speech.

Speech intelligibility was rated by 5 native speakers without either transcription experience or experience working with individuals with speech disorders. Each listener was asked to transcribe the word as well as their confidence in their transcription. The accuracy of the transcriptions across all 5 participants was used to create an intelligibility measure for each individual ranging from 0 (worst) to 100 (best). Participants were categorized into four quartiles defined as very low (0-25%), low (26-50%), middle (51-75%), and high (76-100%). Table 1 summarizes the demographics of the UA-speech participants included in this work.

Table 3: *Features Extracted*

Prosodics	Pitch, Jitter, RMS-Energy, HNR
Spectral	LSF + Δ, MFCC + Δ, Cepstral Peak Prominence (CPP)
TEO	TEO: Amplitude Modulation, Frequency Modulation, Critical Band Areas [20], RMS-Energy, Log-Energy
Glottal	H1-H2 [21], Parabolic Spectrum Parameter [22], Harmonic Richness Factor [23], Glottal Timing Parameters [24]

2.2. Atlanta Motor Speech Disorders Corpus (AMSDC)

The Atlanta Motor Speech Disorders Corpus (AMSDC) contains speech recordings of 99 adults local to the South-Eastern US with acquired neurogenic disorders that resulted in a motor speech disorder [19]. Participants presented with aprosodia, dysarthria, and apraxia of speech. An important distinction in this corpus is the emphasis on regional dialects unique to the southeastern US, especially those of African-American English. Audio recordings were completed in private or semi-private rooms of the Grady Memorial Health System using an AKG C520 condenser microphone with a behind-the-neck band. All participants completed recordings of conversational speech, oral passage readings, sentences with contrasting stress emphasis on specific words, sentences with material sensitive to dysarthria, and affective sentences. Intelligibility was determined by a clinician and 4 graduate speech-language pathology interns based on intelligibility of the read paragraphs.

A subset of 57 participants who presented with dysarthria was selected for analysis in this work. For this work, participants were grouped by the same intelligibility categories used in UA-Speech. An estimated 30 words per participant were segmented from the sentences with contrasting stress emphasis. As few as 12 words were extracted from some participants with low intelligibility and/or connected speech. Table 2 summarizes the clinical and demographic information of the subset of participants included in this work.

3. Methods

3.1. Feature Extraction and Selection

Voiced sections of speech were identified from the word-length recordings of both databases. A variety of prosodic, spectral, Teager Energy Operators (TEO) and glottal features were extracted, and statistics based on those reported in openSmile (i.e., average, min, max, slope) were calculated at the word or sentence level [25]. The full list of features is shown in Table 3, and details regarding the methods for feature calculations are available in descriptions of prior work by the authors [15, 16].

A total of 1595 low-level descriptors (LLD) extracted in MATLAB for each response and normalized across each individual feature by subtracting the global mean and dividing by the global standard deviation. Initial feature reduction involved removing features with a correlation greater than 0.75. Gaussian noise on the order of 10^{-6} was added to the data to ensure the matrix was well conditioned for the machine learning. A 10-fold cross-validation sequential forward feature selection (SFFS) was used to further reduce the size of the feature subsets. In general, the feature selection strategy selected between 3-8 features for each experiment. It is possible by changing the criteria function of the SFFS algorithm, more features would have been included and results could have been improved. However, parameter searches to tune the feature selection algorithm or SVM parameters were not of interest in this paper since the focus was on the impact of cross-dataset setup on classification results. Only the selected features were used to train and test the feature-subset models built using the Support Vector Machine function with a radial-basis function kernel in Matlab. The features in each category of Table 3 were analyzed individually as well as grouped together by category.

3.2. Previous Work for Comparison

Detection of dysarthria from vocal acoustics has occasionally been studied in recent years. Classification results from analysis of vowels by Mekysaka et al. achieved accuracy results ranging from 72-92% depending on the vowel and spoken conditions [26]. DeCicco and Patel used machine learning in vowels from children with dysarthria to automatically detect pitch and duration manipulations for use in alternative and augmentative communication [27]. Vyas et al. extracted MFCCs, skewness, and formants from 40 speech utterances evenly split for training and testing of dysarthria vs controls from UA-Speech [28]. An accuracy of .98 for all speech and .87 for dysarthria speech was reported. Intelligibility classification based on the UA-Speech dataset has been considered by Martinez et al. achieving results of 0.60 for weighted precision and 0.44 for weighted recall in a leave-one-subject-out, 4-class setting [29]. Paja and Falk utilized a two-stage analysis for spastic dysarthria detection and reported results of .83-.95 on the UA-Speech dataset using prosodic, linear-predictive, MFCC, HNR, and Glottal-to-Noise Excitation (GNE) features [6].

3.3. Cross-Database Training and Testing

Previous work on the UA-speech database has been able to take advantage of the singular type of dysarthria presented in the data as well as the uniformity in the type of speech collected. However, in the current work, the aim was to identify features in a broader sense of dysarthria classification for general speech tasks especially considering that the AMSDC contains a variety of types of dysarthria as well as a different emphasis in the collected speech. The cross-database training for this pilot study uses the UA-Speech dataset as the training set since the AMSDC does not contain samples of a control group without dysarthria.

The AMSDC contains continuous speech samples that needed to be manually segmented for this pilot study in making more direct comparisons with UA-Speech. The manual segmentation of the speech from the contrasting stress emphasis portion of the AMSDC resulted in some participants having a maximum of 12 words segmented due to low intelligibility. In order to create an even comparison for cross-database training, all participants were restricted to 12 words for training/testing. Therefore, a reduced-UA-Speech dataset was used with 12 words per person or 336 total words for training purposes. A leave-one-subject-out cross validation strategy was conducted per subject to determine the classification accuracy for the detection of dysarthria presence from the controls in the UA-Speech dataset. Selected classification results from using the reduced-UA-Speech dataset are presented in the first column of Table 4 at the word and participant level. The word level accuracies are calculated

Table 4: C*lassification Accuracy of Predicting Dysarthria, All Models Trained on UA-Speech.*

Feature Type	Reduced UA-Speech	Spastic AMSDC	Spastic/Flaccid AMSDC	All AMSDC
Word-level accuracies				
All	0.661	0.381	0.333	0.329
Prosodics	**0.753**	0.512	0.479	0.548
Glottal	0.711	0.595	0.558	0.554
CPP	0.714	0.548	0.508	0.485
TEO-FM	0.592	0.583	**0.713**	0.706
H1-H2	0.705	0.583	0.604	0.582
Participant-level accuracies				
All	0.750	0.429	0.200	0.193
Prosodics	**0.929**	0.714	0.550	0.684
Glottal	0.857	0.857	0.700	0.649
CPP	0.893	0.571	0.550	0.298
TEO-FM	0.679	0.714	**0.900**	0.877
H1-H2	0.821	0.857	0.800	0.754

by comparing for each word the predicted label of dysarthria or no-dysarthria to the known condition of the individual. The participant-level accuracies are the percentage of participants for which at least 50% of the individual's words were classified correctly. The first column of Table 4 shows the classification accuracy of the training/testing when only the UA-Speech dataset is considered (similar to previous work in Section 3.2). The next columns in Table 4 present results of the classification accuracy when the UA-Speech model is tested on the AMSDC data for spastic, spastic and flaccid, and all combined dysarthria types present in the AMSDC.

4. Results

Table 4 shows that while just over 70% of the words were correctly labeled in the reduced-UA-Speech dataset experiment, the participant level accuracies were much higher. This suggests that the models worked very well on some individuals and very poorly on others. As one example, on the reduced-UA-Speech dataset, the prosodic features classified just 75.3% of the words correctly (0.823 precision, 0.728 recall), but 92.9% of participants were classified correctly. These results suggest a balanced model was built on the reduced-UA-Speech data that did not tend to under- or over-diagnose dysarthria, but did tend to perform poorly on certain individuals.

Previous work by Alghowinem at al. and Tahon et al. are two examples that have used multiple datasets for training/testing in the case of depression or emotions in populations not affected by language or motor disorders [30, 31]. Both studies observed performance declines when a model was trained on one dataset and tested on another, likely due to a variety of factors including recording environment, different speech sample emphasis, etc. Table 4 shows a similar trend with the word- and participant-level accuracies being higher when the training/testing is performed strictly on the UA-Speech data. The training/testing on the UA-Speech alone resulted in the highest word- and participant-level accuracies overall with prosodic features serving as the single best feature category (as highlighted). The overall decline in classification

Table 5: *Correlation Coefficients between Participants' Word-Level Classification Accuracies and Demographic Information within the Various Experiments.*

Feature Type	Gender	Age	Intelligibility Score
Reduced UA-Speech dataset			
All	-0.332	0.093	-0.368
Prosodics	-0.195	-0.191	-0.223
Glottal	-0.304	0.081	**-0.620**
CPP	0.084	-0.022	**-0.533**
TEO-FM	0.003	0.072	**-0.570**
H1-H2	-0.091	0.264	**-0.756**
AMSDC- Spastic Dysarthria only			
All	-0.375	0.331	-0.138
Prosodics	0.444	-0.288	-0.644*
Glottal	0.367	0.066	-0.173
CPP	-0.209	0.101	-0.009
TEO-FM	-0.135	-0.323	-0.049
H1-H2	-0.441	0.226	0.247
AMSDC- All Dysarthria types			
All	0.239	0.216	0.024
Prosodics	**0.532**	0.066	0.124
Glottal	**0.541**	0.132	0.053
CPP	0.158	0.078	0.014
TEO-FM	0.042	-0.040	-0.108
H1-H2	0.206	0.186	-0.056

Bold indicates statistically significant at $p<0.05$, two-tailed

* moderate correlation, not statistically significant as n=7

accuracy shown in the final three columns was expected due to the nature of cross-database training. However, of specific interest was the performance of the TEO-FM, glottal feature group, and the H1-H2 feature categories. While none were able to outperform the overall highest classification accuracy achieved by the prosodics category on the UA-Speech self-training/testing, these features showed the most resilience to the cross training at a participant level. The TEO-FM feature was able to achieve max word- and participant level accuracies of 0.713 and 0.9 when cross tested on the dysarthria of spastic/flaccid types on the AMSDC. This was particularly surprising as the TEO-FM was not a high performing feature for the UA-Speech training model in isolation. It was the only feature tested on the AMSDC to achieve over 0.7 in word accuracy.

A subsequent analysis of the UA-Speech database was conducted to better interpret the results of Table 4. The Pearson correlation coefficients between the word-level classification accuracies of each participant and the demographics provided for the datasets are shown in Table 5. Within the reduced-UA-Speech dataset results, the classification accuracies for the Glottal, CPP, TEO-FM, and H1-H2 feature sets showed a moderately strong negative correlation with the intelligibility score, suggesting that these features classified the best on individuals with more severe dysarthria indicated by a lower intelligibility score. Since it is expected that individuals with more severe dysarthria would have a larger difference in vocal acoustics and speech patterns from those with only mild or no dysarthria, the large negative correlations values are not surprising. While no statistically significant correlations existed in the Spastic-only AMSDC analysis, the moderately high negative correlation of the prosodic feature set to the intelligibility score mimics that of the reduced-UA-Speech dataset, suggesting that the model built on the spastic dysarthria of the UA-Speech dataset worked similarly when applied to the AMSDC spastic dysarthria participants. A result of the all-AMSDC analysis suggested the prosodic and glottal classification accuracies were higher for women than for men; it is unclear why the reduced-UA-Speech dataset model for dysarthria (which was trained on more males than females) would work better on the females than males in the AMSDC dataset. Despite being the best classification performers, there was no significant linear correlation between word accuracies and the glottal, TEO-FM, and H1-H2 features in the AMSDC dataset. However, the overall lack of correlation for the accuracies in the AMSDC is likely due to training on the UA-Speech dataset for the classifier. Table 5 suggests that the intelligibility scores likely played a meaningful role in feature selection from the UA-Speech data as these served as the classifier targets in training.

5. Conclusions and Future Work

This pilot study represents a first attempt to perform cross-database analysis on dysarthria. As expected, the cross-database testing had lower accuracies in general, a finding that supports the challenges of cross-database training and testing. However, the results of this work did discover potentially robust features to cross-database training in the performance of the TEO-FM feature, which has not been a focus of previous work in dysarthria analysis. Participant-level accuracies of the AMSDC tended to be higher when using only the spastic or spastic-flaccid dysarthria instead of all types present in the AMSDC. Previous results have reported higher error rates when attempting to separate dysarthria recordings from participants with similar subtypes of dysarthria (flaccid, spastic, and mixed 26.2%) compared to participants with subtypes known to differ (ataxic vs mixed dysarthria due to amyotrophic lateral sclerosis 11.9%) [32]. The results support multi-staged models trained on single subtypes of dysarthria to maximize performance.

While the words used in training in the UA-Speech database were often multisyllabic, the words segmented from the AMSDC tended be shorter and were not segmented to be phonemically balanced. Work by Martinez et al. informally suggested shorter words were more difficult to classify into dysarthria intelligibility groupings [29]. As segmentation continues, the inclusion of more complex words from other samples in the AMSDC will be investigated to improve alignment with the UA-Speech dataset for increased classification accuracy.

The strong correlations between intelligibility and word-level accuracy on the UA-Speech dataset reflect what is generally expected for this task. However, the fact that this strong correlation did not translate to the AMSDC will need to be further examined. Additionally, the current model represents a binary classification of the presence or absence of dysarthria. Future work will seek to expand the model to include prediction of the dysarthria severity at a word and phrase level.

6. Acknowledgements

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship, Grant No. DGE-1148903. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

7. References

[1] J. R. Duffy, *Motor Speech Disorders: Susbstrates, Differential Diagnosis, and Management*. St. Louis: Mosby, 1995.

[2] H. V. Sharma, "Universal Access: Experiments in Automatic Recognition of Dysarthric Speech," Master of Science, Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 2008.

[3] E. Yilmax, M. Ganzeboom, L. Beijer, C. Cucchiarini, and H. Strik, "A Dutch Dysarthric Speech Database for Individualized Speech Therapy Research," in *Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Portoroz, Slovenia, 2016.

[4] S. R. Shahamiri and S. S. B. Salim, "A Multi-Views Multi-Learners Approach Towards Dysarthric Speech Recognition Using Multi-Nets Artifical Neural Networks," *IEEE Trans. on Neural Systems and Rehabilitation Engineering,* vol. 22, pp. 1053-1063, 2014.

[5] P. Enderby and R. Palmer, *Frenchay Dysarthria Assessment-Second Edition (FDA-2)*. Austin, TX: PRO-ED, Inc;, 2008.

[6] M. S. Paja and T. H. Falk, "Automated Dysarthria Severity Classification for Improved Objective Intelligibility Assessment of Spastic Dysarthric Speech," in *13th Annual Conference of the International Speech Communication Association (INTERSPEECH-2012)*, Portland, OR, 2012, pp. 62-65.

[7] E. C. Guerra and D. F. Lovey, "A Modern Approach to Dysarthria Classification," in *25th Annual International Conference of the IEEE EMBS*, Cancun, Mexico, 2003, pp. 2257-2260.

[8] H. Kim, M. Hasegawa-Johnson, A. Perlman, J. Gunderson, T. Huang, K. Watkin*, et al.*, "Dysarthric Speech Database for Universal Access Research," in *9th Annual Conference of the International Speech Communication Association (INTERSPEECH)*, Brisbane, Australia, 2008, pp. 1741-1744.

[9] X. Menendez-Pidal, J. B. Polikoff, S. M. Peters, J. E. Leonzio, and H. T. Bunnell, "The Nemours Database of Dysarthric Speech," in *Fourth International Conference on Spoken Language (ICSLP-96)*, Philadelphia, PA, 1996, pp. 1962-1965.

[10] F. Rudzicz, A. K. Namasivayam, and T. Wolff, "The TORGO Database of Acoustic and Articulatory Speech from Speaker wtih Dysarthria," *Language Resources and Evaluation,* vol. 46, pp. 523-541, 2012.

[11] J. R. Deller Jr., M. S. Liu, L. J. Ferrier, and P. Robichaud, "The Whitaker Database of Dysarthric (Cerebral Palsy) Speech," *Journal of the Acoustical Society of America,* vol. 93, pp. 3516-3518, 1993.

[12] M. Nicolao, H. Christensen, S. Cunningham, P. Green, and T. Hain, "A Framework for Collecting Realistic Recordings of Dysarthric Speech- the homeService Corpus," in *10th Edition of Language Resources and Evaluation Conference (LREC 2016)*, Portoroz, 2016.

[13] R. Sriranjani, M. Ramasubba Reddy, and S. Umesh, "Improved Acoustic Modeling for Automatic Dysarthric Speech Recognition," in *2015 Twenty First National Conference on Communications (NCC)*, Mumbai, 2015.

[14] I. Laaridh, C. Fredouille, and C. Meunier, "Automatic Detection of Phone-Based Anomalies in Dysathric Speech," *ACM Transactions on Accessible Computing,* vol. 6, p. 9, 2015.

[15] S. Gillespie, E. Moore II, J. S. Laures-Gore, and M. Farina, "Exploratory Analysis of Speech Features Related to Depression in Adults with Aphasia," in *41st IEEE International Conference on Acoustics, Speech, and Signal Processing*, Shanghai, China, 2016, pp. 5185-5189.

[16] S. Gillespie, E. Moore II, J. Laures-Gore, M. Farina, S. Russell, and Y.-Y. Logan, "Detecting Stress and Depression in Adults with Aphasia Through Speech Analysis," in *42nd IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2017)*, New Orleans, LA, USA, 2017.

[17] H. L. Flowers, F. L. Silver, J. Fang, E. Rochon, and R. Martino, "The Incidence, Co-Occurence, and Predictors of Dsyphagia, Dysarthria, and Aphasia After First-Ever Ischemic Stroke," *Journal of Communication Disorders,* vol. 46, pp. 238-248, 2013.

[18] J. R. Orozco-Arroyave, F. Honig, J. D. Arias-Londono, J. F. Vargas-Bonilla, K. Daqrouq, S. Skodda*, et al.*, "Automatic detection of Parkinson's disease in running speech spoken in three different languages," *Journal of the Acoustical Society of America,* vol. 139, pp. 481-500, 2016.

[19] J. S. Laures-Gore, S. Russell, R. Patel, and M. Frankel, "The Atlanta Motor Speech Disorders Corpus: Motivation, Development, and Utility," *Folia Phoniatrica et Logopaedica,* vol. 68, pp. 99-105, 2016.

[20] G. Zhou, J. H. L. Hansen, and J. F. Kaiser, "Nonlinear Feature Based Classification of Speech Under Stress," *IEEE Transactions on Speech and Audio Processing,* vol. 9, pp. 201-216, 2001.

[21] I. R. Titze and J. Sundberg, "Vocal Intensity in Speakers and Singers," *Journal of the Acoustical Society of America,* vol. 91, pp. 2936-2946, 1992.

[22] P. Alku, H. Strik, and E. Vilkman, "Parabolic Spectral Paramater- A New Method for Quanitification of the Glottal Flow," *Speech Communication,* vol. 22, pp. 67-79, 1997.

[23] D. G. Childers and C. K. Lee, "Vocal Quality Factors: Analysis, Synthesis, and Perception," *Journal of the Acoustical Society of America,* vol. 90, pp. 2394-2410, 1991.

[24] J. F. Torres, E. Moore II, and E. Bryant, "A study of glottal waveform features for deceptive speech classification," in *2008 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Las Vegas, NV, USA, 2008, pp. 4489-4492.

[25] F. Eyben, F. Weninger, F. Gross, and B. Schuller, "Recent Developments in OpenSMILE, the Munich Open-Source Multimedia Feature Extractor," in *Proceedings of ACM Multimedia (MM)*, Barcelona, Spain, 2013, pp. 835-838.

[26] J. Mekyska, Z. Smekal, Z. Galaz, Z. Mzourek, I. Rektorova, M. Faundez-Zanuy*, et al.*, "Perceptual Features as Markers of Parkinson's Disease: The Issue of Clinical Interpretability," in *Recent Advances in Nonlinear Speech Processing*, A. Esposito, M. Faundez-Zanuy, A. M. Esposito, G. Cordasco, T. Drugman, J. Solé-Casals*, et al.*, Eds., ed Cham: Springer International Publishing, 2016, pp. 83-91.

[27] T. M. DeCicco and R. Patel, "Machine Classification of Prosodic Control in Dysarthria," *Journal of Medical Speech-Language Pathology,* vol. 18, pp. 35-39, 2010.

[28] G. Vyas, M. K. Dutta, J. Prinosil, and P. Harar, "An Automatic Diagnosis and Assessment of Dysarthric Speech Using Speech Disorder Specific Prosodic Features," in *2016 29th International Conference on Telecommunications and Signal Processing (TSP)*, 2016, pp. 515-518.

[29] D. Martinez, E. Lleida, P. Green, H. Christensen, A. Ortega, and A. Miguel, "Intelligiblity Assessment and Speech Recognizer Word Accuracy Rate Prediction for Dysarthric Speakers in a Factor Analysis Subspace," *ACM Transactions on Accessible Computing,* vol. 6, p. 10, 2015.

[30] S. Alghowinem, R. Goecke, J. Epps, M. Wagner, and J. Cohn, "Cross-Cultural Depression Recognition from Vocal Biomarkers," in *Conference of the International Speech Communication Association (Interspeech-2016)*, San Francisco, 2016.

[31] M. Tahon, M. A. Sehili, and L. Devillers, "Cross-Corpus Experiments on Laughter and Emotion Detection in HRI with Elderly People," in *Social Robotics: 7th International Conference, ICSR 2015, Paris, France, October 26-30, 2015, Proceedings*, A. Tapus, E. André, J.-C. Martin, F. Ferland, and M. Ammi, Eds., ed Cham: Springer International Publishing, 2015, pp. 633-642.

[32] M. V. Mujumdar and R. F. Kubichek, "Design of a Dysarthria Classifier Using Global Statistics of Speech Features," in *2010 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Dallas, Texas, 2010.