Segment-dependent dynamics in predicting Parkinson's disease

James R. Williamson[1], Thomas F. Quatieri [1], Brian S. Helfer[1], Joseph Perricone[1], Satrajit S. Ghosh[2], Gregory Ciccarelli[1], Daryush D. Mehta[1]

[1] MIT Lincoln Laboratory, Lexington, Massachusetts, USA
[2] Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

jrw@ll.mit.edu, quatieri@ll.mit.edu, brian.helfer@ll.mit.edu, joey.perricone@ll.mit.edu, satra@mit.edu, gregory.ciccarelli@ll.mit.edu, daryush.mehta@ll.mit.edu

Abstract

Early, accurate detection of Parkinson's disease may aid in possible intervention and rehabilitation. Thus, simple noninvasive biomarkers are desired for determining severity. In this study, a novel set of acoustic speech biomarkers are introduced and fused with conventional features for predicting clinical assessment of Parkinson's disease. We introduce acoustic biomarkers reflecting the segment dependence of changes in speech production components, motivated by disturbances in underlying neural motor, articulatory, and prosodic brain centers of speech. Such changes occur at phonetic and larger time scales, including multi-scale perturbations in formant frequency and pitch trajectories, in phoneme durations and their frequency of occurrence, and in temporal waveform structure. We also introduce articulatory features based on a neural computational model of speech production, the Directions into Velocities of Articulators (DIVA) model. The database used is from the Interspeech 2015 Computational Paralinguistic Challenge. By fusing conventional and novel speech features, we obtain Spearman correlations between predicted scores and clinical assessments of r = 0.63 on the training set (four-fold cross validation), r = 0.70 on a held-out development set, and r = 0.97 on a held-out test set.

Index Terms: Parkinson's disease, speech biomarkers, phoneme and pause duration, articulatory coordination, neural computational models of motor control

1. Introduction

Parkinson's disease is a neurological disorder with associated progressive decline in motor precision and sensorimotor integration stemming presumably from the basal ganglia. In this disorder, there is a steady loss of cells in the midbrain, leading to speech impairment in nearly 90% of subjects [1]. Speech and voice characteristics of Parkinson's disease include imprecise and incoordinated articulation, monotonous and reduced pitch and loudness, variable speech rate and rushes of breath and pause segments, breathy and harsh voice quality, and changes in intonation and rhythm [2][3][4][5][6].

Early and accurate detection of Parkinson's disease can aid in possible intervention and rehabilitation. Thus, simple noninvasive biomarkers are desired for determining severity of the condition. In this paper, a novel set of speech biomarkers is introduced for predicting the clinical assessment of the Parkinson's disease severity. Specifically, we introduce speech biomarkers representing essential speech production mechanisms of phonation, articulation, and prosody, motivated by changes that are known to occur at different time scales and differentially across speech segments in the underlying neural motor and prosodic brain centers of speech production.

With this motivation, our new features are designed based on two basic tenets: (1) there is a phoneme dependence of the dynamic speech characteristics that is altered in Parkinson's disease, and (2) vocal tract dynamics and stability are altered in Parkinson's disease. We exploit the phoneme dependence of durations and pitch and formant slopes, consistent with findings that certain speech segments are more prone to variation than others in Parkinson's disease [7][8][9]. We also exploit the decline in the precision of articulatory trajectories over different time scales and tasks [10][11].

Our paper is organized as follows. In Section 2, we describe the data collection and preprocessing approaches. In Section 3, we describe our signal models and corresponding signal processing methods for speech feature extraction. Section 4 reports predictor types and prediction results. Section 5 gives conclusions and projections toward future work.

2. Parkinson's database

2.1. Audio recordings

The Parkinson's disease database used in the Interspeech 2015 Computational Paralinguistic Challenge is described in [1]. Assessments of Parkinson's severity are based on the Unified Parkinson's Disease Rating Scale (UPDRS) [12]. The data set is divided into 42 tasks per speaker, yielding 1470 recordings in the training set (35 speakers) and 630 recordings in the development set (15 speakers), both with UPDRS scores provided. It also contains 462 recordings (11 speakers) in the test set, without UPDRS scores provided. The duration of recordings ranges from 0.24 seconds to 154 seconds.

2.2. Audio enhancement

To address noise in the test data, we use an adaptive Wiener-filter approach that preserves the dynamic components of a speech signal while reducing noise [13][14]. The approach uses a measure of spectral change that allows robust

[*] This work is sponsored by the Assistant Secretary of Defense for Research & Engineering under Air Force contract #FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Government. SG was partly supported by a joint collaboration between the McGovern Institute for Brain Research and MIT LL.

and rapid adaptation of the Wiener filter to speech and background events. The approach reduces speech distortion by using time-varying smoothing parameters, with constants selected to produce less temporal smoothing in rapidly-changing regions and greater smoothing in more stationary regions.

2.3. Test set annotation

Identification (ID) labels for 11 test subjects were manually assigned to the monologue, the read-text, and the ten sentence recordings by listening. A 128-component Universal Background Model/Gaussian Mixture Model (UBM/GMM) classifier was used [15], with a feature vector consisting of 16 Mel-frequency cepstral coefficients (MFCCs) and 16 delta-MFCCs, to classify the remaining 330 recordings on the test set. The UBM was trained from all 2,562 recordings in the data set, and subject ID assignments on the test recordings were obtained using 11 subject GMMs, adapted using manually labeled tasks on the test set. Finally, manual correction of 28 of the test subject assignments was done. The final subject ID assignments were not perfect, with number of assigned recordings per subject ranging from 40 to 44.

3. Feature extraction

3.1. Overview of feature sets

Feature development was designed to reflect the three basic aspects of speech production: Phonation (source), articulation (vocal tract), and prosody (intonation and timing). The focus of the features is to characterize variations in dynamics of pitch (phonation), formants (articulation), and rate and waveform (prosody) that reflect Parkinson's severity.

Ten different feature sets are used, designed to characterize changes in speech as a function of Parkinson's severity. Each feature set (FS) comprises a set of raw features, described in Sections 3.2– 3.4. Dimensionality reduction of all FSs is done by z-scoring the raw features and then extracting lower dimensional features using principal components analysis (PCA). The PCA features provide input into regression models that map each FS into a Parkinson's severity (UPDRS) prediction (see Section 4). The FSs, summarized in Table 1, are categorized in terms of three different classes: summary statistics, phoneme dependence, and correlation structure.

Four of the FSs (1, 3, 5, and 7) are effective on short duration tasks, and are applied to all the recordings in the dataset. For these FSs, the data is divided into six different time bins, based on durations of recordings (see Table 2 for details). Each statistical model is trained only on tasks of similar duration. The remaining six FSs are designed to capture longer duration speech dynamics, and are used to characterize changes across subjects when speaking the same sentence. For these FSs, the same three declarative sentences (sentences 2, 4, and 6) are analyzed.

3.2. Summary statistics

FS 1: Delta-MFCC means. Mel-frequency cepstral coefficients (MFCCs) are obtained with openSMILE at a 100 Hz frame rate [16]. Delta-MFCC coefficients are computed using regression over two frames before and after each frame. Then, the mean values of the 16 delta-MFCCs are computed across all frames in each recording.

Table 1. *Feature Set (FS) summary. Three feature types (summary statistics, phoneme dependence, and correlation structure) are applied to time binned and sentence data.*

FS Index	Description	Data Types	# Raw Features	# PCA Features
1	Delta-MFCC means	Time bins	16	15
2	Loudness statistics	Sentences	2	1
3	Phn duration	Time bins	1	1
4	Phn dur. & pitch slope	Sentences	2	2
5	Phn dur. & formant slopes	Time bins	3	2
6	Phn freq.	Sentences	7	1
7	Waveform corr. structure	Time bins	600	6
8	Delta-MFCC corr. structure	Sentences	864 – 912	3
9	Formant corr. structure	Sentences	162 – 171	3
10	Articulatory position corr. structure	Sentences	324 – 342	3

FS 2: Loudness statistics. Loudness is computed from the Perceived Evaluation of Audio Quality (PEAQ) algorithm [17], a psychoacoustic measure of audio quality that has been used for analysis of Lombard speech [18]. Loudness is average energy across critical auditory bands per frame, FS 2 comprises the mean and standard deviation of this feature.

3.3. Phoneme-based features

Changes in phoneme durations and frequencies and in phoneme-dependent pitch and formant slopes reflect the phonemic segment dependence of alterations in phonation and articulation with Parkinson's severity. Phonemic segments are used, along with estimated pitch and formant frequency contours, to generate several phoneme-based feature sets. Using an automatic phoneme recognition algorithm [19], phonemic boundaries are detected, with each segment labeled with one of 40 phoneme classes. The fundamental frequency (pitch) contour is estimated using an autocorrelation method over a 40-ms Hanning window every 1 ms [20]. Formant frequency contours are estimated using a Kalman filter that smoothly tracks the first three spectral modes while also smoothly coasting through non-speech regions [21].

For FS 3, 4, and 5, an aggregation step is performed in which a subset of the 40 phoneme-based measures that are the most highly correlating with Parkinson's severity (on the training set) is linearly combined. In all cases, the top 10 most highly correlating measures are combined using weights w=sign(r)/(1-r^2) [22]. For measures derived from time-bin data, the aggregation is done independently in each time bin. For measures derived from sentences, aggregation is done independently in each sentence.

FS 3: Average phoneme duration. A linear fit is made of the logarithm of pitch over time (within each phonemic segment), yielding a pitch slope (Δlog(Hz)/s) for each phonemic segment. Phoneme durations are then computed for those segments where the pitch slope is marked as valid (i.e., where the absolute pitch slope is less than eight, indicating that the slope is likely derived from a continuous pitch contour).

Average phoneme durations were used originally in classifying depression severity [23].

FS 4: Average phoneme duration and pitch slope. In addition to the average phoneme duration feature, an average log-pitch slope is also used where pitch slopes are valid [22].

FS 5: Average phoneme duration and formant slopes. Linear fits to formant frequencies f_1 and f_2, along with average phoneme durations, are computed from all phoneme segments regardless of pitch slope validity.

FS 6: Phoneme frequencies. The number of occurrences of each phoneme is computed. These counts are normalized to sum to one and sorted from highest to lowest. The top seven of these phoneme frequency measures are used as features.

3.4. Correlation structure features

Measures of the structure of correlations among low-level speech features have previously been applied in the estimation of depression [22][24], the estimation of cognitive performance associated with dementia [25], the detection of changes in cognitive performance associated with mild traumatic brain injury [26], and were first introduced for analysis of EEG signals for epileptic seizure prediction [27].

Channel-delay correlation and covariance matrices are computed from multiple time series. Each matrix contains correlation or covariance coefficients between the channels at multiple time delays. Changes over time in the coupling strengths among the channel signals cause changes in the eigenvalue spectra of the channel-delay matrices. The matrices are computed at four separate time scales, in which successive time delays correspond to different size frame spacings. Overall power (logarithm of the trace) and entropy (logarithm of the determinant) are extracted from the channel-delay covariance matrices at each scale for FSs 8– 10. A detailed description of the correlation structure approach can be found in [27] and its application to speech signals in [22][24].

FS 7: Waveform correlation structure. Correlation structure features are extracted directly from waveform segments, thereby characterizing instability in temporal envelope and phase structure (rhythm and regularity) on different time scales. This technique was applied to each 0.5 s frames with 50% overlap. Each frame is divided into five 0.1 s segments that are treated as separate channels. Utterances ≥ 0.75 s in duration resulted in multiple frames. In these cases the average eigenvalue at each eigenvalue rank is computed across frames. Features are computed at four scales with delay spacings of 3, 7, 15, and 31, with 30 delays per scale. These features reveal an association between Parkinson's severity and reduction in dynamical complexity, as illustrated in Section 3.5.

FS 8: Delta-MFCC correlation structure. Correlation structure features are derived from all 16 delta-MFCCs using four delay scales with spacings 1, 3, 7, and 15, and using 15 delays per scale. Fewer delays are used for the largest scale on sentences 2 and 4 due to their short duration.

FS 9: Formant correlation structure. Correlation structure features are derived from three formant frequencies using the same delay and scale parameters as FS 8.

FS 10: Positions of speech articulators correlation structure. Here, we take advantage of a neurologically plausible, fMRI-validated computational model of speech production, the Directions into Velocities of Articulators (DIVA) model [28]. The DIVA model takes as inputs the first three formants and the fundamental frequency of a speech utterance. Then, through an iterative learning process, the model computes a set of synaptic weights that correspond to different aspects of the speech production process including articulatory commands and auditory and somatosensory feedback errors. We hypothesize that Parkinsonian speech results from impairments along the speech production pathway, and therefore, when the model is trained on Parkinsonian speech, the internal variables will reflect the severity of the disorder. Correlation structure features are derived from the DIVA model's 13 time-varying articulatory position states, are sampled at 200 Hz. The same delay and scale parameters are applied as with FS 8 and FS 9.

3.5. Example of discriminative value

In this section we show the discriminative value of the waveform correlation structure features (FS 7). Figure 1 shows channel-delay correlation matrices obtained from two women with low and high Parkinson's severity (training set files 154 and 303), both speaking the word "crema". The channels are five 0.1 s segments of the audio waveform. The matrix elements are correlation coefficients between channels at different relative time delays. These matrices are obtained at the 3rd time scale, and so successive matrix elements correspond to delays of 15 sub-frames. The matrix from the low severity speaker exhibits more heterogeneity in the correlation patterns, indicating higher waveform complexity.

This difference is quantified using matrix eigenvalues (Figure 2, left panel), ordered largest to smallest. Low UPDRS speech contains greater power in the small eigenvalues. This effect is summarized across multiple speakers by plotting the average eigenvalue for different ranges of Parkinson's severity at each rank. Figure 2 (right panel) shows the eigenvalue averages (in standard units) from all training/development waveforms < 0.5 s in duration for three UPDRS ranges. These averages reveal distinct differences related to Parkinson's severity, even across multiple different speech tasks.

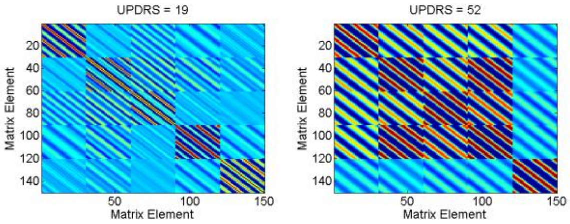

Figure 1: *Correlation matrices from waveform-segment channel-delay matrices for a low (left) and high (right) UPDRS from utterance of "crema".*

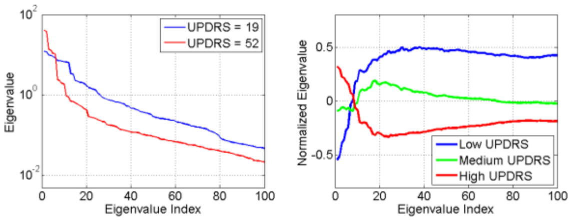

Figure 2: *Left: Eigenspectra for the utterance "crema" from two female speakers with different Parkinson's severity. Right: Average eigenspectra for low (blue), medium (green), and high (red) UPDRS.*

4. Parkinson's score prediction

4.1. Gaussian staircase regression models

There is a separate predictor for each FS, and the predictor outputs are linearly combined to produce the system's net UPDRS prediction. Each predictor uses a Gaussian staircase (GS) regression model [22][24], which comprises an ensemble of six Gaussian classifiers, each trained on data with different UPDRS partitions between Class 1 (lower UPDRS) and Class 2 (higher UPDRS). The Class 1 partitions are 0– 15, 0– 24, 0– 33, 0– 42, 0– 51, and 0– 60, and the Class 2 partitions are the complement of these partitions. Additional regularization of the densities is obtained by adding the constant 4 to the diagonal elements of the data-normalized covariance matrices. The GS output score is the ratio of the log of the summed likelihoods from each ensemble of Gaussians. This is followed by applying a 2nd-order univariate regression model, created from the GS training scores, to the GS test score to generate a UPDRS prediction.

4.2. Within-subject prediction averaging

For each FS, subject ID labels are the basis for computing, across all tasks for that subject, a weighted prediction average based on the tasks processed by each FS. The provided subject IDs are used on the training/development sets, and the subject IDs from our procedure (Section 2.3) are used on the test set.

Time-bin predictions. For FS 1, 3, 5 and 7, there are six different regression models, one for each time bin. The utility of each FS in each time bin is assessed based on the Spearman correlation of its predictions on the training set (using 4-fold cross validation). Within each FS and subject ID, the predictor outputs are combined across time bins ID using weights $w = r^2/(1\text{-}r^2)$ if $r > 0$; $w = 0$ otherwise. Table 2 lists the normalized weights for each FS, with weights summing to one in each column. Observe that the FSs accumulate most of their evidence from short duration tasks.

Table 2. *Weights used for combining predictions across recording duration bins for feature sets (FS).*

Time bin range (s)	Time bin weights			
	FS 1	FS 3	FS 5	FS 7
0.0 – 0.5	0.54	0.00	0.00	0.32
0.5 – 0.75	0.03	0.44	0.56	0.00
0.75 – 2.0	0.23	0.56	0.11	0.03
2.0 – 4.0	0.20	0.00	0.01	0.58
4.0 – 7.0	0.00	0.00	0.00	0.01
7.0 –	0.00	0.00	0.32	0.06

Sentence predictions. The remaining FSs use data from matched sentences (sentences 2, 4 and 6), which range in duration from 1.56 s to 9.63 s. The sentence tasks allow for phonemic timing and correlation structure biomarkers that have been used to predict various neurological states [11][12][21][22][23]. Within each FS and test subject ID, the GS training and test scores are averaged across the three sentences. A 2nd-order univariate regression model obtains a sentence-based prediction for each FS and test subject.

4.3. Fusing predictors

The ten FS predictors are applied individually to the training set (four-fold cross-validation with held out speakers) and to the development set (Table 3). We used linear combinations of the predictors with three different weight vectors w_1, w_2, and w_3. With w_1, each predictor is weighted equally. To improve performance, we also applied differential weighting of the ten predictors. To choose the weights, we conducted a grid search over all 1,024 possible weight combinations in which each weight can have a value of 1 or 2. From this search, we obtained two different weight vectors based on different constraints. The first weight vector is w_2, the weights that yield the highest average of 1) Spearman correlation on the training set, 2) Spearman correlation on the development set, and 3) a test metric score. The second weight vector is w_3, the weights that yield the highest test metric score. The test metric is the Pearson correlation between previous submission scores and the Spearman correlations between the prediction vectors that had produced the previous scores and a candidate prediction vector. The test metric equals one if a candidate prediction vector has Spearman correlation of one with the true UPDRS scores. To compute this metric, we used ten previously obtained scores, which range between 0.12 and 0.89. These scores are our previous eight submissions and two Baseline results (Table 3, rows 1, 3 of [29]). After training on the combined train and development sets, we submitted test results using w_2 and w_3, obtaining Spearman correlations of $r = 0.96$ and $r = 0.97$, respectively.

Table 3. *Performance for each feature set predictor and for linear combinations of predictors.*

Predictor	Train *r*	Devel. *r*	w_1	w_2	w_3
1	0.59	0.39	1	2	1
2	0.29	0.35	1	2	2
3	0.56	0.67	1	1	1
4	0.24	0.37	1	1	1
5	0.46	0.10	1	2	1
6	0.10	0.09	1	1	1
7	0.56	0.53	1	1	2
8	0.43	0.49	1	2	1
9	0.46	0.21	1	1	2
10	0.39	0.25	1	1	1
w_1	0.61	0.75	–	–	–
w_2	0.62	0.78	–	–	–
w_3	0.63	0.70	–	–	–

5. Conclusions and discussion

We applied standard and novel speech features predicting levels of Parkinson's disease. Our features capture segment-based dynamics across phonemes, formant frequencies and articulatory positions, based on an understanding of the effect of Parkinson's disease on speech production components. Our ongoing work involves the further enhancement of current features and establishment of new features, with emphasis on approaches that provide a neurological basis for understanding the effect of Parkinson's disease on speech.

6. Acknowledgements

The authors thank Dr. Elizabeth Godoy and Dr. Christopher Smalt for providing code and guidance on the extraction of loudness features, and Dr. Douglas Sturim for helpful discussions on automatic speaker identification.

7. References

[1] J. Orozco-Arroyave, J. Arias-Londono, J. Vargas-Bonilla, M. González-Rátiva, and E. Nöth, "New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease," in *Proceedings of the 9th Language Resources and Evaluation Conference (LREC)*, 2014, pp. 342–347.

[2] G. J. Canter, "Speech characteristics of patients with Parkinson's disease: I. Intensity, pitch, and duration," *Journal of Speech and Hearing Disorders*, vol. 28, no. 3, pp. 221–229, 1963.

[3] G. J. Canter, "Speech characteristics of patients with Parkinson's disease: II. Physiological support for speech." *Journal of Speech and Hearing Disorders*, vol. 30, no. 1, pp. 44–49, 1965.

[4] G. J. Canter, "Speech characteristics of patients with Parkinson's disease: III. Articulation, diadochokinesis, and over-all speech adequacy," *Journal of Speech and Hearing Disorders*, vol. 30, no. 3, pp. 217–224, 1965.

[5] J. A. Logemann, H. B. Fisher, B. Boshes, and E. R. Blonsky, "Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients," *Journal of Speech and Hearing Disorders*, vol. 43, no. 1, pp. 47–57, 1978.

[6] J. E. Sussman and K. Tjaden, "Perceptual measures of speech from individuals with Parkinson's disease and multiple sclerosis: Intelligibility and beyond," *Journal of Speech, Language, and Hearing Research*, vol. 55, no. 4, pp. 1208–1219, 2012.

[7] J. A. Logemann and H. B. Fisher, "Vocal tract control in Parkinson's disease," *Journal of Speech and Hearing Disorders*, vol. 46, no. 4, pp. 348–352, 1981.

[8] S. Skodda, and U. Schlegel, "Speech rate and rhythm in Parkinson's disease," *Movement Disorders*, vol. 23, no. 7, pp. 985–992, 2008.

[9] S. Skodda, "Aspects of speech rate and regularity in Parkinson's disease," *Journal of the neurological sciences*, vol. 310, no. 1–2, pp. 231–236, 2011.

[10] S. G. Hoberman, "Speech techniques in aphasia and Parkinsonism," *Journal - Michigan State Medical Society*, vol. 57, no. 12, pp. 1720–1723, 1958.

[11] I. J. Rusz, R. Cmejla, T. Tykalova, H. Ruzickova, J. Klempir, V. Majerova, J. Picmausova, J. Roth, and E. Ruzicka, "Imprecise vowel articulation as a potential early marker of Parkinson's disease: Effect of speaking task," *The Journal of the Acoustical Society of America*, vol. 134, no. 3, pp. 2171–2181, 2013.

[12] C. G. Goetz, B. C. Tilley, S. R. Shaftman, G. T. Stebbins, S. Fahn, P. Martinez-Martin, ... and N. LaPelle, "Movement Disorder Society–sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results," *Movement Disorders*, vol. 23, no. 15, pp. 2129–2170, 2008.

[13] T. F. Quatieri and R. B. Dunn, "Speech enhancement based on auditory spectral change," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2002, pp. I-257.

[14] T. F. Quatieri and R. A. Baxter, "Noise reduction based on spectral change," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 1997.

[15] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital signal processing*, vol. 10, no. 1, pp. 19–41, 2000.

[16] 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.

[17] T. Thiede, W. C. Treurniet, R. Bitto, C. Schmidmer, T. Sporer, J. G. Beerends, and C. Colomes, "PEAQ-The ITU standard for objective measurement of perceived audio quality," *Journal of the Audio Engineering Society*, vol. 48, no. 1/2, pp. 3–29, 2000.

[18] E. Godoy and Y. Stylianou, "Unsupervised acoustic analyses of normal and Lombard speech, with spectral envelope transformation to improve intelligibility," in *13th Annual Conference of the International Speech Communication Association, September 9–13, Portland, Oregon, Proceedings*, 2012, pp. 1472–1475.

[19] W. Shen, C. White, T. J. Hazen, "A comparison of query-by-example methods for spoken term detection," in *Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing*, 2010.

[20] P. Boersma and D. Weenink, "Praat, a system for doing phonetics by computer," 2001.

[21] D. D. Mehta, D. Rudoy, and P. J. Wolfe, "Kalman-based autoregressive moving average modeling and inference for formant and antiformant tracking," *The Journal of the Acoustical Society of America*, vol. 132, no. 3, pp. 1732–1746, 2012.

[22] 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 ACM International Workshop on Audio/Visual Emotion Challenge (AVEC)*, 2014, pp. 65–72.

[23] A. Trevino, T. F. Quatieri, and N. Malyska, "Phonologically-based biomarkers for major depressive disorder," *EURASIP Journal on Advances in Signal Processing*, vol. 42, pp. 1–18, 2011.

[24] 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.

[25] B. Yu, T. F. Quatieri, J. W. Williamson, and J. Mundt, "Prediction of cognitive performance in an animal fluency task based on rate and articulatory markers," in *15th Annual Conference of the International Speech Communication Association, September 9–13, Portland, Oregon, Proceedings*, 2014.

[26] B. S. Helfer, T. F. Quatieri, J. R. Williamson, L. Keyes, B. Evans, W. N. Greene, J. Palmer, and K. Heaton, "Articulatory dynamics and coordination in classifying cognitive change with preclinical mTBI," in *15th Annual Conference of the International Speech Communication Association, September 9–13, Portland, Oregon, Proceedings*, 2014.

[27] J. R. Williamson, D. Bliss, D. W. Browne, and J. T. Narayanan, "Seizure prediction using EEG spatiotemporal correlation structure," *Epilepsy and Behavior*, vol. 25, no. 2, pp. 230–238, 2012.

[28] F. H. Guenther, S. S. Ghosh, and J. A. Tourville, "Neural modeling and imaging of the cortical interactions underlying syllable production," *Brain and Language*, vol. 96, no. 3, pp. 280–301, 2006.

[29] B. Schuller, S. Steidl, A. Batliner, S. Hantke, F. Hönig, J. R. Orozco-Arroyave, E. Nöth, Y. Zhang, F. Weninger, "The INTERSPEECH 2015 Computational Paralinguistics Challenge: Nativeness, Parkinson's & Eating Condition," in *INTERSPEECH 2015 – 16th Annual Conference of the International Speech Communication Association, September 6–10, Dresden, Germany, Proceedings*, 2015.