



# Acoustic Indicators of Speech Motor Coordination in Adults With and Without Traumatic Brain Injury

Tanya Talkar<sup>1,2</sup>, Nancy Pearl Solomon<sup>3,4</sup>, Douglas S. Brungart<sup>3,4</sup>, Stefanie E. Kuchinsky<sup>3</sup>, Megan M. Eitel<sup>3,5,6</sup>, Sara M. Lippa<sup>3,7</sup>, Tracey A. Brickell<sup>3,4,6,7,8,9</sup>, Louis M. French<sup>3,4,6,7</sup>, Rael T. Lange<sup>3,6,7,8,9,10</sup>, Thomas F. Quatieri<sup>1,2</sup>

<sup>1</sup>Speech and Hearing Bioscience and Technology, Harvard University, Boston, MA, USA

<sup>2</sup>MIT Lincoln Laboratory, Lexington, MA, USA

<sup>3</sup>Walter Reed National Military Medical Center, Bethesda, MD, USA

<sup>4</sup>Uniformed Services University of the Health Sciences, Bethesda, MD, USA

<sup>5</sup>Henry M. Jackson Foundation, Rockville, MD, USA

<sup>6</sup>Traumatic Brain Injury Center of Excellence, Silver Spring, MD, USA

<sup>7</sup>National Intrepid Center of Excellence, Bethesda, MD, USA

<sup>8</sup>Centre of Excellence on Post-traumatic Stress Disorder, Ottawa, ON, Canada

<sup>9</sup>General Dynamics Information Technology, Falls Church, VA, USA

<sup>10</sup>University of British Columbia, Vancouver, BC, Canada

ttalkar@g.harvard.edu, nancy.p.solomon.civ@mail.mil, douglas.s.brungart.civ@mail.mil, stefanie.e.kuchinsky.civ@mail.mil, megan.m.eitel.ctr@mail.mil, sara.m.lippa.civ@mail.mil, tbrickell.dvbic@gmail.com, louis.m.french.civ@mail.mil, rael.lange@gmail.com, quatieri@ll.mit.edu

## Abstract

A traumatic brain injury (TBI) can lead to various long-term effects on memory, attention, and mood, as well as the occurrence of headaches, speech, and hearing problems. There is a need to better understand the long-term effects of a TBI for objective tracking of an individual's recovery, which could be used to determine intervention trajectories. This study utilizes acoustic features derived from recordings of speech tasks completed by active-duty service members and veterans (SMVs) enrolled in the Defense and Veterans Brain Injury (DVBIC)/Traumatic Brain Injury Center of Excellence (TBICoE) 15-Year Longitudinal TBI Study. We hypothesize that the individuals diagnosed with moderate to severe TBI would demonstrate motor speech impairments through decreased coordination of the speech production subsystems as compared to individuals with no history of TBI. Speech motor coordination is measured through correlations of acoustic feature time series representing speech subsystems. Eigenspectra derived from these correlations are utilized in machine learning models to discriminate between the two groups. The fusion of correlation features derived from the recordings achieves an AUC of 0.78. This suggests that residual motor impairments from moderate to severe TBI could be detectable through objective measures of speech motor coordination.

**Index Terms:** speech motor coordination, acoustic analysis, traumatic brain injury, machine learning

## 1. Introduction

The Traumatic Brain Injury Center of Excellence (TBICoE) has reported approximately 430,720 incidents of traumatic brain injuries (TBIs) among U.S. service members between 2000-2020 [1]. While the majority of the TBI cases seen at Veterans Affairs hospitals are mild TBI (mTBI), there are many cases of moderate and severe TBI, which can affect an individual's quality of

life in the long term. These chronic effects may lead to impaired memory, attention, and motor functions [2]. To mitigate these long-term effects, many individuals will undergo clinical treatments targeting the relevant impairments. Individuals also present with a wide variety of motor speech impairments, and may benefit from speech therapy [3]. Quantitative analysis of speech production in individuals who have had at least a moderate TBI could lead to better understanding of how TBI affects speech production, as well as measures that could be used for planning intervention trajectories and tracking progress.

Acoustic features, as well as vocal biomarkers derived from these features, have been used to characterize and detect speech production differences in individuals with a history of TBI [4, 5, 6, 7]. In a previous investigation of a cohort of participants from the same larger study used here, it was found that speech during diadochokinetic (DDK) tasks (e.g. rapid repetition of syllables) was slower in adults with moderate through severe TBI [4]. In other previous work, correlation structures formed from acoustic features, used as a proxy measure of speech motor coordination, have also shown promise in detecting changes in cognitive performance for high school athletes with preclinical mTBI [6]. In addition, correlation structures derived from read speech, free speech prompts, and DDK tasks have been used to detect and characterize lingering motor impairments in mTBI patients who had been clinically considered recovered [7]. In particular, features derived from the correlation structures are used to capture the coordination both within and across speech production subsystems using representative acoustic features from the articulatory, laryngeal, and respiratory subsystems [8].

In this paper, we posit that the application of speech motor coordination features to individuals who have a history of at least moderate TBI will provide further insight into the ways that TBI affects speech motor coordination. In section 2, we describe how we apply our correlation structure analysis to acoustic features derived from a dataset of speech recordings col-

lected at Walter Reed National Military Medical Center (WRNMMC). Eigenvalues extracted from this analysis are used to create a gaussian mixture model (GMM) classifier to discriminate between individuals who have a history of moderate through severe TBI versus controls. The eigenvalues are further used to characterize speech motor coordination across and within three main speech production subsystems - articulatory, laryngeal, and respiratory. Section 3 describes the results from these analyses. Section 4 discusses relevant interpretations from our results and details our plan for further analysis with this dataset.

## 2. Methods

### 2.1. Participants

Speech and neuropsychological testing was carried out at WRNMMC as part of the DVBC-TBICoE 15-year Longitudinal TBI Study (Sec721 NDAA FY2007). Additional details on inclusion criteria, group definition, and recruiting procedures are found in Lange et al. (2019). The current analyses included 116 participants from a possible pool of 213 participants with complete speech data. Eleven participants were excluded due to having an equivocal or unknown TBI history. Individuals were also excluded because of invalid scores on performance validity tests ( $n = 22$ ) [9, 10]. Finally, participants were excluded if they had history of a mild TBI ( $n = 64$ ). TBI severity was classified as follows: Moderate TBI: loss of consciousness (LOC) >30 mins-24 hours, post-traumatic amnesia (PTA) 1-7 days, and intracranial abnormality (ICA) present or absent; Severe TBI: LOC >24 hours, PTA >7 days, and ICA present or absent; Penetrating TBI: breach of the cranial vault and/or dura mater by external object (e.g., bullet, shrapnel) and/or skull fragment (i.e., skull fracture). In total, 36 individuals with a history of moderate (31%), severe (38%), or penetrating (31%) TBI were included. Individuals with no history of TBI included 37 injured controls (orthopedic/soft tissue injury with no evidence of alteration of consciousness (AOC), LOC, or PTA as result of injury) and 43 non-injured controls, for a total of 80 controls.

### 2.2. Procedure

Speech-language assessments took place in a double-walled sound-attenuating booth with the participant seated in front of a video monitor and the examiner visible through the examination window. Recordings were collected onto a laboratory computer via a cardioid dynamic microphone (Shure PG48) with a constant mouth-to-microphone distance of 14 cm. Recordings were made through an internal soundcard (RME Hammerfall DSP Multiface II) using a software audio recorder function (MATLAB 2007). Participants were instructed via a pre-recorded video to repeat one trial each of repetition of four diadochokinetic (DDK) sequences: /pʌ/, /tʌ/, /kʌ/, /pʌtəkʌ/, as fast and as accurately as possible on a single breath. Participants subsequently watched a wordless picture story and were prompted to retell the story for their free-speech sample. If participants were unable to provide an adequate story-telling sample, they were asked to tell the examiner about their day. Participants were then asked to read The Caterpillar passage [11].

### 2.3. Low-level feature extraction

Acoustic feature time series and delta time series were extracted from each recording. These features have been selected to represent the three speech subsystems of interest (articulatory, laryngeal, respiratory). The first three vocal tract resonances (F1-

F3) were calculated at 100Hz using the Kalman-based autoregressive moving average (KARMA) software tool, which provides a continuous time series of vocal tract resonances using an energy-based voice detector, utilizing a Kalman smoother to estimate formants through silent gaps in the signal [12]. Fundamental frequency (F0), mel-frequency cepstral coefficients (MFCCs), and harmonic-to-noise ratio (HNR) were extracted using the Praat software at 1000 Hz, 200 Hz, and 100 Hz respectively [13, 14]. Cepstral peak prominence (CPP) and creak, representing acoustic correlates of voice quality, were extracted using custom MATLAB scripts at 100 Hz [15, 16, 17]. Intensity, or the speech envelope, was extracted at 100 Hz using a custom MATLAB script that provides a smooth contour of amplitude peaks based on an iterative time-domain signal envelope estimation [18]. This algorithm estimates both the contributions of the respiratory system and resonance-harmonics interaction to the amplitude modulation of a speech envelope.

A software-based acoustic-to-articulatory inversion algorithm was used to extract out tract variables (TVs), representing the movement of articulators in the vocal tract [19]. The algorithm takes in an acoustic speech signal and outputs the trajectories of six different vocal tract constriction variables, from the Task Dynamics model: lip aperture (LA), lip protrusion (LP), tongue body constriction location (TBCL) and degree (TBCD), and tongue tip constriction location (TTCL) and degree (TTCD). Each of these was extracted from recordings and used as low-level time series. The articulators were used as a comparison to formants, which are a higher-level representation of the articulators.

### 2.4. High-level feature extraction

Multivariate auto- and cross-correlations of the low-level features were calculated to produce proxy measures of coordination within and across the underlying speech motor subsystems [20, 21]. Channel-delay correlation matrices were calculated from combinations of the acoustic and articulatory low-level feature time series, representing the time-delay embedding space by expanding the dimensionality of the low-level acoustic and articulatory time series. This process is described in Figure 1. To capture the relative feature coupling strengths across multiple time delays, these correlation matrices were calculated at four delay scales, with delay spacings of 1, 3, 7, and 15 data samples for all of the acoustic feature time series.

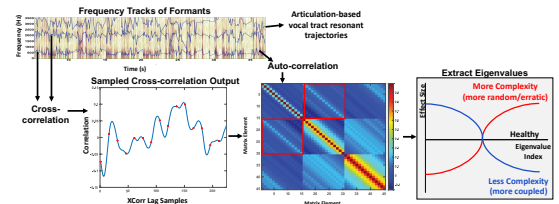


Figure 1: Exemplar process of high-level correlation structure extraction using F1, F2, and F3 time series, and subsequent interpretation of eigenvalue patterns.

Correlations across features were calculated within and across speech subsystems to capture the interaction of these subsystems during speech production. Feature combinations (e.g. feature 1 x feature 2) were computed by concatenating the feature vectors. For example, F0 x formants yielded a  $4 * t$  feature matrix, where  $t$  is the total number of samples in the time series, due to a  $1 * t$  F0 matrix concatenated with a  $3 * t$  formant

matrix. This feature matrix was used to calculate the resulting correlation matrices. Correlation matrices for speech tasks were calculated for the following feature combinations and their delta-time series equivalents: F0, formants, envelope, MFCC, TV, envelope x formants, F0 x formants, F0 x envelope, F0 x envelope x formants. Additionally, correlations were computed for vocal source features: HNR, CPP, creak, F0 x CPP, F0 x HNR, F0 x HNR x CPP, F0 x envelope x HNR x CPP, envelope x HNR x CPP, envelope x CPP. This led to a total of 27 different feature combinations. Any feature correlated with F0 was interpolated to a sampling rate of 1000 Hz using spline interpolation. A masking technique was used to include only voiced segments for all correlations, using the locations where F0 > 0.

Eigenvalues of all resulting correlation matrices were extracted by rank-order (from greatest to smallest). Each correlation matrix correlating  $n$  time series had  $15 * n$  eigenvalues extracted to comprise the eigenspectrum (e.g., correlations of F0 and envelope lead to  $15 * 2 = 30$  eigenvalues). Eigenvalues from each delay scale for a single task and feature combination were concatenated to form a final feature vector with  $n * 15 * 4$  elements to input into a classifier. The eigenvalues from each delay scale were also used to characterize the complexity of the signals, using the interpretation outlined in Figure 1.

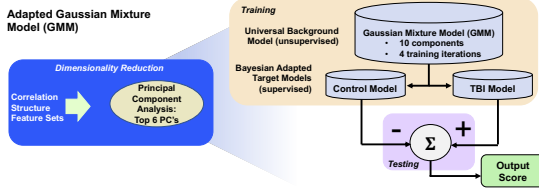


Figure 2: *Gaussian Mixture Model (GMM) architecture used to discriminate between participants with or without a history of TBI.*

## 2.5. Classification

Classification of participants with or without a history of TBI was done with leave-one-out cross validation using a GMM (Figure 2). For each task and feature combination, the top 6 principal components from principal component analysis (PCA) were extracted from the eigenspectra vector. In each cross-validation fold, an ensemble of ten GMMs were created over four iterations using the PCA features from the training set. Control and TBI GMMs were generated using supervised adaptation of the ensemble GMMs, a technique that is commonly used in speaker verification [22]. The likelihood of the held-out test participant belonging to each group was computed by summing up the likelihood of the participant belonging to each GMM in the ensemble. The final model output score was the ratio of the log likelihood of the participant belonging to the TBI GMMs over the log likelihood of the participant belonging to the control GMMs. The model output scores were used to compare model performance across features and tasks using the area under the receiver operating characteristic curve (AUC), created by plotting the true positive rate against the false positive rate at various threshold settings.

Features were fused if they produced sufficient cross-validation accuracy within each training fold. For each participant, the model output scores and training set AUC were saved from each feature and task combination. First, fusion was computed across all features within a task through the process out-

lined in eq. (1), where  $p$  is a specific participant,  $t$  is a specific task, and  $f$  is a specific feature.

$$score_{p,t} = \frac{\sum_{f \in Features} mask_{p,t,f} * score_{p,t,f}}{\sum_{f \in Features} mask_{p,t,f}} \quad (1)$$

For each participant and feature, a mask was created where the value was 1 if the associated training AUC was greater than 0.5. A sum was taken across all model output scores from features that met that criterion and divided by the number of features used. The fusion score across multiple tasks was computed by taking the sum of the model output scores for each participant, and dividing by the number of tasks. This was first done to combine information from the DDK tasks, and then to combine scores from DDK, read, and free speech. These fused scores, across a single task and across all tasks, were subsequently used to calculate the fused AUC.

## 3. Results

### 3.1. Classification

Fusing across all speech tasks and features resulted in an AUC of 0.78. AUCs for individual tasks varied, but, notably, the AUC for the read speech task was 0.73 and the AUC for the DDK task was 0.72. Table 1 lists the AUC values for fusions within a task. Table 1 also includes a comparison of the individual AUCs for formants and TVs.  $AUC > 0.595$  has a p-value  $< 0.05$ .

Table 1: *Classification performance (AUC) across tasks for all fused features, formants, and TVs.*

Task	All Features	Formants	TVs
Diadochokinetic (DDK)	$0.72 \pm 0.11$	0.54	0.52
Read Speech	$0.73 \pm 0.10$	0.66	0.70
Free Speech	$0.67 \pm 0.11$	0.41	0.50
Fused All Tasks	<b>0.78</b>		

### 3.2. Speech motor coordination characterization

Cohen's d effect sizes of the eigenvalues derived from TBI vs control groups were calculated across all features and tasks. Figure 3 shows the effect sizes for the correlation of F0 across all tasks, and the interaction between F0 and formants (delay scale: 15 ms; 15 samples), two features which showed up as top individual features. Figure 4 shows the effect sizes for the correlation of formants and tract variables (delay scale: 150 ms; 15 samples), to compare across articulatory measures. The eigenvalues are ranked from largest to smallest, and an effect size  $> 0$  indicates that the eigenvalue was greater in the TBI group. As shown in Figure 1, a pattern of positive effect sizes for larger eigenvalues and negative effect sizes for smaller eigenvalues would suggest lower complexity of signals in the TBI group, while the opposite pattern would suggest higher complexity of signals in the TBI group.

The patterns in the left panel of Figure 3 indicate that participants with a history of TBI have lower complexity of F0 across all tasks, suggesting more coupled vocal source movements. The morphology of the patterns for the interaction of F0 and formants (right panel of Figure 3) also suggest that there is lower complexity in the movements of TBI participants for all tasks, indicating higher coupling between the two subsystems as well. The eigenvalue patterns for the formants and TVs

also suggest that talkers with TBI have lower complexity (Figure 4), but this was only true for the read speech and DDK tasks. The effect sizes for the coordination features of formants and TVs for the free speech task were generally small across all the eigenvalue indexes, suggesting that other features may have been driving the classification performance for that speech task.

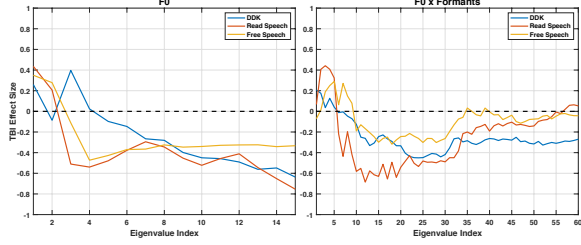


Figure 3: *Cohen's d* effect sizes of eigenspectra extracted from coordination of F0 (left) and the interaction of F0 and formants (right). Eigenvalues are ranked from largest to smallest.

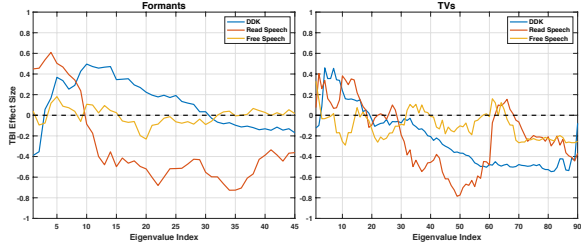


Figure 4: *Cohen's d* effect sizes of eigenspectra extracted from coordination of formants (left) and tract variables (right).

## 4. Discussion

In this paper, we describe a speech protocol and analytical methodology to detect and characterize speech motor coordination in individuals who have had a history of moderate through severe TBI. A GMM created from a fusion of correlation features derived from acoustic time series was able to discriminate between control and TBI participants with an AUC of 0.78. These results provide evidence that utilization of coordination features derived from speech tasks can provide insight into speech motor coordination issues in individuals with a history of TBI, and highlights the residual effects that can be present in an individual who has experienced at least moderate TBI. It suggests that this approach could be used to augment existing clinical assessments to monitor long-term recovery for active-duty service members and veterans.

Comparison of eigenvalue patterns across tasks provided a characterization of the speech of individuals with a history of TBI and suggests that articulators are more coupled during read speech and DDK tasks in participants with this. Patterns suggest that TBI participants had higher coupling within the laryngeal speech production subsystem as compared to controls. Patterns also implied higher coupling of the interaction between the articulatory and laryngeal subsystems during all tasks for the TBI group. This highlights that it is important to look across subsystems for characterization of speech motor coordination, as speech production relies on precise timings within and across speech subsystems. Classification performance differed across the three tasks, specifically with features derived

from read speech and DDK tasks performing better than features derived from free speech. This suggests that there may be task dependent speech motor coordination demands in chronic TBI, which leads to coordination differences being witnessed in some tasks, but not others [23]. As the study progresses, future work will analyze how the results generalize to other speech tasks, whether the speech patterns are consistent across longitudinal speech samples collected from the same talkers, and whether these patterns will persist in a larger group of participants.

In previous studies of major depressive disorder (MDD), coordination features derived from tract variables performed just as well as, and sometimes better than, coordination features derived from formants in discriminating between individuals with MDD vs neurotypical controls [24, 25]. We saw similar results discriminating between the two groups in this paper. We aim to analyze the movements that are derived through acoustic-to-articulatory methods and quantify how they relate to the information presented by formants. We hope that this will enable a better understanding of the coordination of the movements of articulators.

The cohort of individuals in this dataset, both individuals with a history of TBI and controls, have also been assessed for additional comorbidities, such as post traumatic stress disorder, depression, mood disorders, and sleep disorders. As with the current paper comparing individuals with and without TBI, MDD can manifest as low complexity in speech motor coordination [20, 26]. Previous studies have primarily focused on individuals who have a single diagnosis, without documented comorbidities. In subsequent analysis, we plan to analyze the effects of additional diagnoses in the control and TBI groups to see if we can better characterize and cluster speech motor coordination differences for individuals accordingly. Objective and quantitative analysis of these differences will provide insight into how the different conditions affect speech production subsystems, and could lead to better tracking and personalized intervention trajectory planning by clinicians.

## 5. Acknowledgements

DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited. This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering. The views expressed in this manuscript are those of the authors and do not necessarily represent the official policy or position of the Defense Health Agency, Department of Defense, or any other U.S. government agency. This work was prepared under Contract HT0014-19-C-0004 with DHA Contracting Office (CO-NCR) HT0014 and, therefore, is defined as U.S. Government work under Title 17 U.S.C.101. Per Title 17 U.S.C.105, copyright protection is not available for any work of the U.S. Government. For more information, please contact dha.TBICOEinfo@mail.mil. UNCLASSIFIED The identification of specific products or scientific instrumentation is considered an integral part of the scientific endeavor and does not constitute endorsement or implied endorsement on the part of the authors, U.S. Department of Defense, or any component agency. This work was supported by the Traumatic Brain Injury Center of Excellence.



## 6. References

- [1] "DoD TBI worldwide numbers." [Online]. Available: <https://www.health.mil/About-MHS/OASDHA/Defense-Health-Agency/Research-and-Development/Traumatic-Brain-Injury-Center-of-Excellence/DoD-TBI-Worldwide-Numbers>
- [2] J. L. Ponsford, M. G. Downing, J. Olver, M. Ponsford, R. Acher, M. Carty, and G. Spitz, "Longitudinal follow-up of patients with traumatic brain injury: outcome at two, five, and ten years post-injury," *Journal of neurotrauma*, vol. 31, no. 1, pp. 64–77, 2014.
- [3] P. Gandhi, S. Tobin, M. Vongphakdi, A. Copley, and K. Watter, "A scoping review of interventions for adults with dysarthria following traumatic brain injury," *Brain injury*, vol. 34, no. 4, pp. 466–479, 2020.
- [4] N. P. Solomon, D. S. Brungart, J. R. Wince, J. C. Abramowitz, M. M. Eitel, J. Cohen, S. M. Lippa, T. A. Brickell, L. M. French, and R. T. Lange, "Syllabic diadochokinesis in adults with and without traumatic brain injury: Severity, stability, and speech considerations," *American Journal of Speech-Language Pathology*, 2021.
- [5] A. C. Lammert, J. R. Williamson, A. Hess, T. Patel, T. F. Quatieri, H. J. Liao, A. Lin, and K. J. Heaton, "Noninvasive estimation of cognitive status in mild traumatic brain injury using speech production and facial expression," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 2017, pp. 105–110.
- [6] B. S. Helfer, T. F. Quatieri, J. R. Williamson, L. Keyes, B. Evans, W. N. Greene, T. Vian, J. Lacirignola, T. Shenk, T. Talavage *et al.*, "Articulatory dynamics and coordination in classifying cognitive change with preclinical mTBI," in *Fifteenth Annual Conference of the International Speech Communication Association*, 2014.
- [7] T. Talkar, S. Yuditskaya, J. R. Williamson, A. C. Lammert, H. Rao, D. Hannon, A. O'Brien, G. Vergara-Diaz, R. Delaura, D. Sturim, and *et al.*, "Detection of subclinical mild traumatic brain injury (mtbi) through speech and gait," in *Interspeech 2020*, 2020.
- [8] T. Talkar, J. R. Williamson, D. J. Hannon, H. M. Rao, S. Yuditskaya, K. T. Claypool, D. Sturim, L. Nowinski, H. Saro, C. Stamm *et al.*, "Assessment of speech and fine motor coordination in children with autism spectrum disorder," *IEEE Access*, vol. 8, pp. 127 535–127 545, 2020.
- [9] R. T. Lange, S. M. Lippa, L. M. French, J. M. Bailie, R. L. Gartner, A. E. Driscoll, M. M. Wright, J. K. Sullivan, N. V. Varbedian, E. A. Barnhart *et al.*, "Long-term neurobehavioural symptom reporting following mild, moderate, severe, and penetrating traumatic brain injury in us military service members," *Neuropsychological rehabilitation*, 2019.
- [10] S. M. Lippa, L. M. French, R. S. Bell, T. A. Brickell, and R. T. Lange, "United states military service members demonstrate substantial and heterogeneous long-term neuropsychological dysfunction after moderate, severe, and penetrating traumatic brain injury," *Journal of neurotrauma*, vol. 37, no. 4, pp. 608–617, 2020.
- [11] R. Patel, K. Connaghan, D. Franco, E. Edsall, D. Forgit, L. Olsen, L. Ramage, E. Tyler, and S. Russell, "the caterpillar: A novel reading passage for assessment of motor speech disorders," 2013.
- [12] 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.
- [13] P. Boersma and D. Weenink, "Praat: doing phonetics by computer, version 6.0.37." 2018. [Online]. Available: <http://www.praat.org/>
- [14] P. Boersma, "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," in *Proceedings of the institute of phonetic sciences*, vol. 17, no. 1193. Citeseer, 1993, pp. 97–110.
- [15] J. Hillenbrand, R. A. Cleveland, and R. L. Erickson, "Acoustic correlates of breathy vocal quality," *Journal of Speech, Language, and Hearing Research*, vol. 37, no. 4, pp. 769–778, 1994.
- [16] Y. D. Heman-Ackah, D. D. Michael, and G. S. Goding Jr, "The relationship between cepstral peak prominence and selected parameters of dysphonia," *Journal of Voice*, vol. 16, no. 1, pp. 20–27, 2002.
- [17] R. Fraile and J. I. Godino-Llorente, "Cepstral peak prominence: A comprehensive analysis," *Biomedical Signal Processing and Control*, vol. 14, pp. 42–54, 2014.
- [18] R. L. Horwitz-Martin, T. F. Quatieri, E. Godoy, and J. R. Williamson, "A vocal modulation model with application to predicting depression severity," in *2016 IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. IEEE, 2016, pp. 247–253.
- [19] G. Sivaraman, V. Mitra, H. Nam, M. Tiede, and C. Espy-Wilson, "Unsupervised speaker adaptation for speaker independent acoustic to articulatory speech inversion," *The Journal of the Acoustical Society of America*, vol. 146, no. 1, pp. 316–329, 2019.
- [20] J. R. Williamson, D. Young, A. A. Nierenberg, J. Niemi, B. S. Helfer, and T. F. Quatieri, "Tracking depression severity from audio and video based on speech articulatory coordination," *Computer Speech & Language*, vol. 55, pp. 40–56, 2019.
- [21] T. F. Quatieri, J. R. Williamson, C. J. Smalt, J. Perricone, T. Patel, L. Brattain, B. Helfer, D. Mehta, J. Palmer, K. Heaton *et al.*, "Multimodal biomarkers to discriminate cognitive state," *The role of technology in clinical neuropsychology*, vol. 409, 2017.
- [22] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted gaussian mixture models," *Digital signal processing*, vol. 10, no. 1-3, pp. 19–41, 2000.
- [23] W. Ziegler, "Speech motor control is task-specific: Evidence from dysarthria and apraxia of speech," *Aphasiology*, vol. 17, no. 1, pp. 3–36, 2003.
- [24] C. Y. Espy-Wilson, A. C. Lammert, N. Seneviratne, and T. F. Quatieri, "Assessing neuromotor coordination in depression using inverted vocal tract variables," in *Interspeech*, 2019, pp. 1448–1452.
- [25] N. Seneviratne, J. R. Williamson, A. C. Lammert, T. F. Quatieri, and C. Espy-Wilson, "Extended study on the use of vocal tract variables to quantify neuromotor coordination in depression," in *Proc. Interspeech*, vol. 2020, 2020.
- [26] J. R. Williamson, T. F. Quatieri, B. S. Helfer, G. Ciccirelli, 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.