INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

Vocal Biomarker Assessment Following Pediatric Traumatic Brain Injury: A Retrospective Cohort Study

Camille Noufi[1,2], Adam C. Lammert[1], Daryush D. Mehta[1], James R. Williamson[1], Gregory Ciccarelli[1], Douglas Sturim[1], Jordan R. Green[3], Thomas F. Campbell[4], Thomas F. Quatieri[1]

[1]MIT Lincoln Laboratory, Lexington, Massachusetts, USA
[2]Stanford University, Stanford, California, USA
[3]Mass General Hospital, Boston, Massachusetts, USA
[4]University of Texas, Dallas, Texas, USA

cnoufi@stanford.edu, [adam.lammert, daryush.mehta, jrw, gregory.ciccarelli, sturim]@ll.mit.edu, jgreen2@mghihp.edu, thomas.f.campbell@utdallas.edu, quatieri@ll.mit.edu

Abstract

Recommendations following pediatric traumatic brain injury (TBI) support the integration of instrumental measurement to aid perceptual assessment in recovery and treatment plans. A comprehensive set of sensitive, robust and non-invasive measurements is therefore essential in assessing variations in speech characteristics over time following pediatric TBI. In this paper, we discuss a method for measuring changes in the speech patterns of a pediatric cohort of ten subjects diagnosed with severe TBI. We apply a diverse set of both well-known and novel feature measurements to child speech recorded throughout the year following diagnosis. We analyze these features individually and by speech subsystem for each subject as well as for the entire cohort. In children older than 72 months, we find highly significant ($p < 0.01$) increases in pitch variation and number of unique phonemes spoken, shortened pause length, and steadying articulation rate variability. Younger children exhibit similar steadied rate variability alongside an increase in articulation complexity. Nearly all speech features significantly change ($p < 0.05$) for the cohort as a whole, confirming that acoustic measures expanding upon perceptual assessment are needed to identify efficacious treatment targets for speech therapy following TBI.[1]

Index Terms: pediatric, traumatic brain injury, vocal biomarkers, dysarthria, dysphonia, speech subsystems, acoustics, longitudinal, retrospective, cohort study

1. Introduction

The recovery process following severe traumatic brain injury (TBI) in children varies widely. Each individual may experience a combination of cognitive, perceptive, emotional and motor function impairment [1]. Voice and speech disorders commonly occurring post-TBI indicate impairment to the brain's widely distributed speech network. Commonly reported characteristics of speech altered by dysphonia, a laryngeal function disorder, and dysarthria, a motor-speech disorder, are found within the articulatory, prosodic and phonatory speech subsystems [1–4]. A closer look at these affected speech subsystems can provide refined cues to impairments of the brain's speech network as well as identify efficacious treatment targets for speech therapy [4].

Recommendations following pediatric TBI support the integrated use of objective, instrumental measurement alongside perceptual assessment for informing and aiding in recovery and treatment plans [5]. Vocal biomarkers used to monitor and predict cognitively-impaired speech have been studied for several decades [4] and many commercial acoustic analysis programs are available [2]. However, to our knowledge, acoustic analysis has only been applied to the longitudinal study of speaking rate following pediatric TBI [6]. In this article, we define and apply a diverse set of acoustic, audio-based vocal biomarkers to study TBI-affected speech of a pediatric population over time. We apply a diverse set of both well-known and novel speech feature measurements to speech recordings taken monthly over the course of one year and summarize the longitudinal changes in these features via a trend profile. This allows us to track how particular characteristics of each speech subsystem are changing in relation to each other. Section 2 details this methodology. In Section 3, we analyze the trends exhibited by the cohort and comment on the nuanced relationship between recovery trends and age. In Section 4, we discuss the applicability of this feature set in assisting assessment throughout recovery. Section 5 concludes our work.

2. Methods

2.1. Study Design and Data Collection

We apply our feature analysis to speech data originally collected by Campbell *et al.* to understand longitudinal changes in the perceptual 'Percent Consonants Correct - Revised' (PCC-R) intelligibility measure. This data consists of recorded conversations between a subject and a trained examiner [7]. Due to the large amount of preprocessing required to utilize this dataset for acoustic analysis, we focus on ten out of fifty-six subjects within the original study. The ten-subject cohort was selected to be representative of the age range within the original sample, include both males and females, and have corresponding recordings of good perceptual quality.

Each subject had been recently diagnosed with a severe

[1]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 author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering. This research was supported, in part, by the National Institute on Deafness and other Communication Disorders Grant R01DC0368.

http://dx.doi.org/10.21437/Interspeech.2019-1200

Table 1: *Acoustic measurements and their related speech characteristic (CO=Consonant, VW=Vowel)*

Measurement	Speech Characteristic
Speech vs. Articulation Rate	Intraphrase Pauses
CO and VW Durations	Articulation Rate
Variance of CO and VW Durations	Rate variability
CO / VW Occurrence	Complexity & Precision
Number of Unique Phonemes	Complexity & Precision
F0 Mean	Abnormal Pitch
F0 Variance	Monotonous Pitch

TBI. Recorded speech sampling sessions began as soon as possible after the child was able to produce at least 10 intelligible words. Available metadata for each subject consisted of gender, age in months at injury, and age in months at the first speech sampling session. The cohort consisted of six males and four females ranging from 43 months to 123 months in age at the first session.

2.2. Data Preprocessing

Several preprocessing steps were necessary to measure signal-based changes to a child's speech within and across sessions. The raw audio consisted of 15 to 40 minutes of conversational speech between two to four speakers (occasionally family members were present during a session). Because the speech was originally obtained for perceptual assessment, steps to reduce channel effects and noise were not needed during the recording process. Therefore, speaker segmentation and channel-noise reduction were applied to all recordings.

Speaker separation was performed via a Gaussian-mixture-model-based diarization tool using 100 ms minimum duration segments and 2 speaker clusters [8, 9]. However, diarization models are rarely trained on child speech, so manual tuning was performed before and after automatic diarization to ensure the maximum amount of a subject's speech was extracted while excluding other speakers, lengthy silences and non-stationary noise. This was done by selecting the timestamps of child speech detected by the diarization tool, padding each segment with an additional 0.2 seconds to retain lengthy consonant utterances, and then fusing together segments less than 1 second apart. Segments less than 2 seconds in length were discarded after this step, as they consisted of fragmented speech that would confound prosody and timing measurements in downstream processing. Each file was then manually checked for any remaining cross-talk or other perceivable non-child-speech elements. The intact child speech was enhanced using the Optimal Modified Minimum Mean-Square Error Log-Spectral Amplitude method [10] to remove additive non-stationary noise while limiting distortion of clean speech. We were able to extract $(\mu, \sigma) = (2.14, 1.23)$ minutes of clean, connected child speech from each session recording.

2.3. Feature Extraction

The feature set was selected based on commonly reported perceptual speech characteristics affected by TBI in both longitudinal [6, 7] and case-control studies [1, 2, 5, 11, 12] as well as features that have captured similar speech degradation due to diagnosed dyarthria and dysphonia [4], amyotrophic lateral sclerosis [4, 13], depression [14–17], and Parkinson's disease [4].

2.3.1. Formant-Track Coordination

Many studies report that TBI may degrade the physiological coordination underlying vocal tract trajectories and reduce the magnitude and complexity of articulator movement during speech [1, 2, 4, 11]. Vocal tract resonance trajectories are captured via formant frequency tracks. We use multivariate auto- and cross-correlations as a proxy measure of coordination between these formant-tracks. Changes over time in the coupling strengths among the formant tracks cause changes in the eigenvalue spectra of the resulting correlation matrices; weakly coupled formant-tracks may indicate more complex interactions between the articulators. Williamson *et al.* first applied this multivariate correlation approach to epileptic seizure prediction from multichannel EEG [18] and subsequently to the tracking and prediction of major depressive disorder from audio-based vocal signals [14, 19].

The first three formant-tracks (F1-F3) are estimated using the extended Kalman algorithm by Mehta and Rudoy [20]. The algorithm uses a 10 ms frame interval and is seeded with average age- and gender-dependent formant values reported by Lee *et al.* [21]. Estimates of the F3 above a threshold of 4.5k Hz are truncated. The tracks are then subdivided into 20 second segments.

For each 20 second segment, a channel-delay correlation matrix is computed from the three low-level formant tracks. Each matrix contains correlation coefficients between the tracks at a relative time delay. Five matrices are computed at five delay scales (10, 30, 70, 100 and 150 ms) with 15 time-delays used per scale. The eigenvalues of a matrix are extracted by rank-order. For each delay scale, the resulting eigenvalue spectra of the 20 second segments are averaged to obtain the mean eigenvalue decomposition for the session. This approach is performed per session, per subject. Williamson *et al.* provide a complete mathematical description of this method in [19].

2.3.2. Phoneme-based Prosody and Timing

We measure prosody at the phoneme level under the hypothesis that this captures nuanced information in motor and linguistic planning, timing and execution [3, 4, 12, 17]. The KALDI ASpIRE Chain Model is used to extract all uttered phonemes and silences from a session recording and tag them with predicted English phoneme type and duration. We perform a statistical summary across the phoneme collection to obtain the mean, median, variance, inter-quartile range, and count of all 39 possible Standard American English phonemes per session. We combine these statistics to create the timing and prosody measurements in Table 1. Speaking rate includes speech pauses while articulation rate measures phonemes only.

2.3.3. Phonation

Cepstral Peak Prominence (CPP) and the H2-H1 relative amplitude difference (H2H1) have been shown to quantify breathy [22, 23] and creaky phonation [22, 24], respectively. Breathiness shows low periodicity and steep spectral slope in the source spectrum, its cepstral representation containing peaks with smaller amplitudes. A lower CPP is therefore an indication of a more breathy voice [22, 23]. The source spectrum of a creaky voice contains a much larger second harmonic (H2) amplitude than first harmonic (H1) amplitude. A large H2H1 indicates a flatter spectral tilt and more creaky phonation due to glottal constriction [22, 24].

We measure CPP and H2H1 at each 10ms frame of a sin-

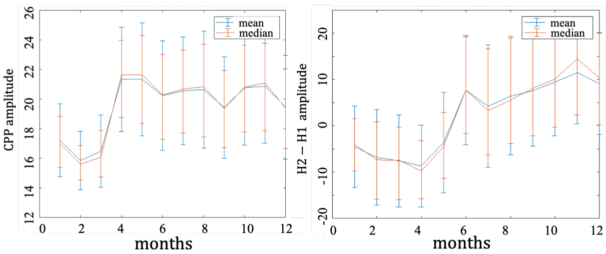

Figure 1: *Cepstral Peak Prominence (left) and H2-H1 amplitude difference (right) of a 10-year old child over one year.*

gle session recording. We then perform a statistical summary within the session to obtain the mean, median, skew, variance, and inter-quartile range as our quantitative representation of breathy and creaky voice quality for the session. Figure 1 depicts a change in CPP and H2H1 over time for a single subject. To measure pitch-related abnormalities, we extract fundamental frequency F0 estimations via Praat at each 10ms frame. Summary statistics of F0 are calculated for each recording [4].

2.4. Trend Summarization

Longitudinal changes to these features are measured via linear correlation with time. These correlations summarize the direction and strength of change to a speech feature between the initial and final sampling sessions.

Formant-track coordination complexity is summarized over time by examining how the eigenvalue spectrum of a correlation matrix changes over the course of 12 months. The average eigenvalue spectrum of each session is linearly correlated per-rank with session number (1 is the first session, 12 is the final session, approximately 12 months later) to obtain a vector describing this change over time. Equation 1 defines this vector for a single subject at one delay scale.

$$(\rho_\lambda[r], p_\lambda[r]) = corr(\lambda_r, S)\ \forall r \in \mathbb{Z}^{1:(M\times N)} \tag{1}$$

where λ_r is the eigenvalue of a particular session at rank r, M is the number of channels, N is the number of delays per channel, $S \in \mathbb{Z}^{1\times i}$ is the vector containing all sessions, and i is the number of sessions. ρ_λ is the rank-ordered vector describing the formant-track correlation with and $\mathbf{p}$ is the p-value significance vector.

When low-rank eigenvalues negatively correlate with time and higher-rank eigenvalues positively correlate with time, information is distributing throughout the correlation matrix as recovery time increases, representing formant track interactions becoming more ‘complex’.

Phoneme-based and phonatory changes over time are summarized via linear correlation of the summary statistics with session number. An individual subject’s summary statistics are correlated with his/her specific S vector (see 1). Age-independent group trends are calculated by first normalizing each subject’s measurements across all sessions to zero mean and unit standard deviation (e.g. $zscore(CPP_{mean})$) to remove effects of age-influenced baseline values. These normalized measurements are then concatenated. S vectors for each subject are concatenated as well. For each speech characteristic c_n, a Pearson correlation is performed between these two aggregate vectors to obtain correlation coefficient ρ_n and significance p-value p_n. These age-independent group trends are calculated for all ten subjects, for younger children (less than 72 months old), and for older children (72 months and older).

To capture developmental influence on baseline measurements and trends, age-dependent group correlations are calculated by replacing each subject’s S vector with his/her age (in months) at each session and skipping normalization.

We group this set of features by speech subsystem and produce a trend profile as a comprehensive summarization. Change in formant-track complexity is represented by the correlation of the first-order eigenvalue with session number $\rho_\lambda[1]$. Correlation coefficients and significance values of mean and variance summarize the change to phoneme-based and phonation features. The trend profile can be used to record the proposed set of quantitative measurements periodically, or be used as a comprehensive summarization chart. Figure 2 shows a trend profile of an individual subject. Here, ρ and p indicating significant directional trends are represented with $+$ and $-$.

Category	Group	Feature	Measure	Trend
Motor Coordination and Timing	Rate and Prosody	Average Rate	Speech	+ +
			Articulation	+
		CO	Rate	+
			Rate Variability	- -
		VW	Rate	+
			Rate Variability	- -
		Pitch (F0)	Average	NT
			Variation	-
	Articulation	Coordination		+
		CO/VW Ratio	Duration	NT
			Occurrence	+
		Number of Unique Phonemes		+
		PCC-R		NT
Voice Quality	Phonation	Breathy Voice	Avg	+
			Var	-
		Creaky Voice	Avg	NT
			Var	NT

Figure 2: *Trend Profile of youngest subject. Symbols indicate direction and strength of speech changes over one year:* $\rho > 0.5$*,* $\rho > 0.3$ *are denoted by* $++$*,* $+$*, respectively.* $\rho < -0.5$*,* $\rho < -0.3$ *are denoted by* $--$*,* $-$*, respectively.* $|\rho| < 0.3$ *and/or* $p > 0.05$ *are denoted by NT, representing "no trend".*

3. Results

Figure 2 provides a case analysis of the youngest member of the cohort. This particular child increases her speaking rate more so than articulation rate, indicating a shortening of pause length over time and suggesting an improvement in linguistic processing speed as hypothesized by Campbell et al [6]. Her consonant rate steadies over time, suggesting an increasing in timing control [4, 5]. The moderate increase in complexity of formant interaction is consistent with moderate increases in consonant/vowel occurrence and number of unique phonemes produced per session, indicating a trend toward more complex articulation ability coupled with an improvement in linguistic processing ability.

Table 2 quantitatively summarizes the trend results for the whole cohort. Across older children, we find statistically significant increases in speaking rate, pitch variation and number of unique phonemes spoken, while we see significant reduction in variability of consonant duration. Across subjects younger children, we find similar reduction in variability of consonant duration, while also seeing an increase in articulatory coordination. Both younger and older children demonstrate movement away from breathy voice into more typical or creaky voice. Via analysis of the whole cohort, we conjecture that some of the directional trends lacking significance in the age subgroups might be due to their small sample size. For example, number of unique phonemes increases moderately but significantly ($p << 0.01$), while consonant duration decreases moderately and significantly ($p = 0.01$).

We correlate each feature measurement with age (skipping

Table 2: *Trend profile of cohort using Pearson correlation to indicate direction and strength of speech changes over one year. YP = younger children, OP = older children. Bold values indicate measures of high significance with* $p \leq 0.01$.

Articulation		**YP**	**OP**	**All**	**With Age**
Formant Coord.		**ρ = 0.50**	0.05	0.22	0.05
		p = 0.01	>0.1	>0.1	0.6
Consonant/ Vowel Ratio	Duration	-0.052	0.013	0.043	0.547
		0.728	0.917	0.65	0
	Occ.	0.253	0.167	0.193	0.404
		0.086	0.174	0.039	0
# Unique Phns		0.241	**0.438**	**0.357**	0.322
		0.102	**0**	**0**	0.103
Prosody		**YP**	**OP**	**All**	**With Age**
Overall Rate	Speech	0.329	**0.513**	**0.437**	0.568
		0.024	**0**	**0**	0
	Artic.	0.196	0.202	0.199	0.618
		0.188	0.099	0.033	0
Consonant	Duration	-0.261	-0.225	**-0.240**	-0.495
		0.076	0.065	**0.010**	0
	Variation	**-0.505**	**-0.390**	**-0.437**	-0.311
		0	**0.001**	**0**	0.001
Vowel	Duration	-0.183	-0.213	-0.201	-0.604
		0.218	0.081	0.032	0
	Variation	-0.318	-0.259	**-0.283**	-0.42
		0.029	0.033	**0.002**	0
Pitch	Mean	0.325	**-0.692**	-0.325	-0.776
		0.026	**0**	0.026	0
	Variation	-0.301	**0.538**	-0.301	-0.688
		0.04	**0**	0.039	0
Phonation		**YP**	**OP**	**All**	**With Age**
Breathy Voice	Mean	-0.355	**-0.317**	**-0.306**	0.455
		0.014	**0.009**	**0.001**	0
	Variation	0.207	0.232	0.222	-0.395
		0.163	0.057	0.017	0
Creaky Voice	Mean	**0.471**	**0.469**	0.102	0.423
		0.001	**0**	0.271	0
	Skew	**-0.418**	**-0.319**	-0.015	-0.315
		0.003	**0.008**	0.871	0.001

the normalization step described in Section 2.4) to examine influence of age on baseline measurements and trends. Table 2 suggests that age at injury significantly influences measurement values at the first session and, in some cases, may indicate the direction of change one would see in normal development. For example, change in normalized articulation rate is not significant for either age group when normalized. However, we see a strong correlation between increasing articulation rate and increasing age. This suggest that, for this cohort, baseline articulation rates are higher for older subjects. Negative correlations between age and fundamental frequency are the strongest, indicating pitch is closely linked with age. Converse examples are weaker correlations between age and the change in number of unique phonemes (in older children) and consonant rate variability, both highly significant when normalized. These weaker correlations with age may indicate that changes to these features are more closely tied to recovery than to normal development.

4. Discussion

The results presented in Section 3 indicate moderate, significant directional speech changes distributed across the articulatory, phonatory and prosodic subsystems. The changes seen in formant-track coordination, unique phonemes uttered, and phone duration require improvement of motor coordination and precision, each in a slightly different way. These changes are more dramatic and consistent amongst children older than 72 months. The steadying of phone duration variability in conjunction with an increase in speech rate and a shortening of speech pause length are highly significant across the cohort, and are consistent with findings in [4, 12, 17].

Several features that may correlate significantly with recovery also correlated significantly with age. Thus, future work providing age-specific normative measurements (as in [5,7,25].) or conducting longitudinal case-control studies (as in [6]) would aid in judging the efficacy of age-specific trajectories. For example, formant-track complexity increases in the younger children, while the older children are able to produce a more diverse set of sounds after a year. Both indicate an increase in motor coordination abilities but are different manifestations, perhaps affected by other developmental traits. (Expanded analyses on the implications of age can be found in [26].)

It is important to note the shortcomings of this particular method applied to this specific dataset. The trends seen exhibited by this small cohort may not generalize to a larger population and, based on the existing literature, are likely not suitable stand-alone predictors of the wide-ranging symptoms and recovery patterns children exhibit following a TBI. Additionally, this dataset is unique in that it provided an opportunity to apply acoustic features longitudinally, but it was not originally designed for such acoustic analysis. Therefore, the feature set was constrained to be as robust as possible to unrestrained, spontaneous child speech of variable length; it is by no means an exhaustive set [4]. Finally, the utilized diarization and segmentation tools used are trained on healthy adult speech, and hence likely perform worse on child or disordered speech. Although performance errors may be consistent within a subject, research into how automatic signal analysis tools perform on atypical speech would help create more reliable and inclusive assessment systems.

5. Conclusion

This study aimed to track the acoustic changes within and across speech subsystems of children who had suffered a severe TBI. The results support existing evidence that the brain's widely distributed speech network is often impacted following a TBI, and features measuring the nuanced, outward manifestations of linguistic processing and motor control may be viable indicators of trauma to the brain. This analysis shows promise for tracking an individual's speech changes over time following TBI and provides a baseline approach for future longitudinal studies utilizing vocal biomarkers.

6. Acknowledgements

We thank Drs. Michael Brandstein, Bengt Borgstrom and Malcolm Slaney for their contribution to and feedback on this work.

7. References

[1] B. E. Murdoch and D. G. Theodoros, *Traumatic Brain Injury: Associated Speech, Language, and Swallowing Disorders*. San Diego: Singular Thomson Learning, 2001.

[2] B. E. Murdoch, F. M. Lewis, and C. Knuepffer, "Assessment and treatment of speech and language disorders in traumatic brain injury," in *Brain Injury Medicine, 2nd Edition: Principles and Practice*, N. D. Zasler, D. I. Katz, , and R. D. Zafonte, Eds. New York: Demos Medical Publishing, 2013, ch. 65, pp. 1086–1110.

[3] W. Ziegler and D. von Cramon, "Spastic dysarthria after acquired brain injury: An acoustic study," *International Journal of Language & Communication Disorders*, vol. 21, no. 2, pp. 173–187, 1986.

[4] E. C. Thompson-Ward and D. G. Theodoros, "Acoustic analysis of dysarthric speech," in *Dysarthria: A Physiological Approach to Assessment and Treatment*, B. E. Murdoch, Ed. Cheltenham, UK: Stanley Thornes Ltd, 1998, ch. 4, p. 102129.

[5] S. McCauley, E. Wilde, V. Anderson, G. Bedell, S. Beers, T. Campbell, S. Chapman, L. Ewing-Cobbs, J. Gerring, G. Gioia, H. Levin, L. Michaud, M. Prasad, B. Swaine, L. Turkstra, S. Wade, and K. Yeates, "Recommendations for the use of common outcome measures in pediatric traumatic brain injury research," *Journal of Neurotrauma*, vol. 29, no. 4, pp. 678–705, 2012.

[6] T. F. Campbell and C. A. Dollaghan, "Speaking rate, articulatory speed, and linguistic processing in children and adolescents with severe traumatic brain injury," *Journal of Speech, Language, and Hearing Research*, vol. 38, no. 4, pp. 864–875, 1995.

[7] T. F. Campbell, C. Dollaghan, J. Janosky, H. Rusiewicz, S. Small, F. Dick, J. Vick, and P. Adelson, "Consonant accuracy after severe pediatric traumatic brain injury: A prospective cohort study," *Journal of Speech, Language, and Hearing Research*, vol. 56, no. 3, pp. 200–209, 2013.

[8] F. C. Douglas Reynolds, Patrick Kenny, "A study of new approaches to speaker diarization," in *INTERSPEECH 2009 – 10th Annual Conference of the International Speech Communication Association, Brighton, UK, Proceedings*, 2009, pp. 1047–1050.

[9] E. C. D. R. Stephen Shum, Najim Dehak and J. Glass, "Exploiting intra-conversation variability for speaker diarization," in *INTERSPEECH 2011 – 12th Annual Conference of the International Speech Communication Association, Florence, Italy, Proceedings*, 2011, pp. 945–948.

[10] I. Cohen and B. Berdugo, "Speech enhancement for nonstationary noise environments," *Signal Processing*, vol. 81, no. 11, pp. 2403 – 2418, 2001.

[11] B. S. Helfer, T. F. Quatieri, J. R. Williamson, L. Keyes, B. Evans, W. N. Greene, T. Vian, J. Lacirignola, T. Shenk, T. Talavage, J. Palmer, and K. Heaton, "Articulatory dynamics and coordination in classifying cognitive change with preclinical mtbi," in *INTERSPEECH 2014 – 15th Annual Conference of the International Speech Communication Association, September 14-18, Singapore, Proceedings*, 2014, pp. 485–489.

[12] M. Falcone, N. Yadav, C. Poellabauer, and P. Flynn, "Using isolated vowel sounds for classification of mild traumatic brain injury," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 7577–7581.

[13] J. R. Green, Y. Yunusova, M. S. Kuruvilla, J. Wang, G. L. Pattee, L. Synhorst, L. Zinman, and J. D. Berry, "Bulbar and speech motor assessment in als: Challenges and future directions," *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, vol. 14, no. 7-8, pp. 494–500, 2013.

[14] J. R. Williamson, T. F. Quatieri, B. S. Helfer, G. Ciccarrelli, 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*, ser. AVEC '14, 2014, pp. 65–72.

[15] S. Alghowinem, R. Goecke, M. Wagner, J. Epps, M. Breakspear, and G. Parker, "Detecting depression: A comparison between spontaneous and read speech," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*.

[16] B. S. Helfer, T. F. Quatieri, J. R. Williamson, D. D. Mehta, R. Horwitz, and B. Yu, "Classification of depression state based on articulatory precision," in *INTERSPEECH 2013 – 14th Annual Conference of the International Speech Communication Association, Lyon, France, Proceedings*, 2013.

[17] A. C. Trevino, T. F. Quatieri, and N. Malyska, "Phonologically-based biomarkers for major depressive disorder," *EURASIP Journal on Advances in Signal Processing*, no. 1, p. 42, Aug 2011.

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

[19] 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.

[20] D. D. Mehta and D. Rudoy, "Kalman-based autoregressive moving average modeling and inference for formant and antiformant tracking," *The Journal of the Acoustical Society of America*, vol. 135, no. 5, 2014.

[21] S. Lee, A. Potamianos, and S. Narayanan, "Acoustics of childrens speech: Developmental changes of temporal and spectral parameters," *The Journal of the Acoustical Society of America*, vol. 105, no. 3, pp. 1455–1468, 1999.

[22] 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.

[23] Y. D. Heman-Ackah, D. D. Michael, and G. S. Goding, "The relationship between cepstral peak prominence and selected parameters of dysphonia," *Journal of Voice*, vol. 16, no. 3, pp. 20–27, 2002.

[24] P. Keating, M. Garellek, and J. Kreiman, "Acoustic properties of different kinds of creaky voice," in *18th International Congress of Phonetic Sciences, Glascow, Scotland, UK*, 2015.

[25] J. Coplan and J. R. Gleason, "Unclear speech: Recognition and significance of unintelligible speech in preschool children," *Pediatrics*, vol. 82, no. 3, 1988.

[26] C. Noufi, A. C. Lammert, D. D. Mehta, J. R. Williamson, G. Ciccarelli, D. Sturim, J. R. Green, T. F. Campbell, and T. F. Quatieri, "Tracking acoustic speech changes over time following pediatric traumatic brain injury," *IEEE Journal of Selected Topics in Signal Processing: Special issue on automatic assessment of health disorders based on voice, speech and language processing. Submitted 15 May 2019, Under Review.*