Articulatory Dynamics and Coordination in Classifying Cognitive Change with Preclinical mTBI

Brian S. Helfer1, Thomas F. Quatieri1, James R. Williamson1
Laurel Keyes1, Benjamin Evans1, W. Nicholas Greene1, Trina Vian1, Joseph Lacirignola1,
Trey Shenk2, Thomas Talavage2, Jeff Palmer1, Kristin Heaton3

1MIT Lincoln Laboratory, Lexington, Massachusetts, USA
2Purdue University, West Lafayette, Indiana, USA
3USARIEM, Natick, MA, USA

Abstract

Speech analysis has shown potential for identifying neurological impairment. With brain trauma, changes in brain structure or connectivity may result in changes in source, prosodic, or articulatory aspects of voice. In this work, we examine articulatory components of speech reflected in formant tracks, and how changes in track dynamics and coordination map to cognitive decline. We address a population of athletes regularly receiving impacts to the head and showing signs of preclinical mild traumatic brain injury (mTBI), a state indicated by impaired cognitive performance occurring prior to concussion. We hypothesize that this preclinical damage results in 1) changes in average vocal tract dynamics measured by formant frequencies, their velocities, and acceleration, and 2) changes in articulatory coordination measured by a novel formant-frequency cross-correlation characterization. These features allow machine learning algorithms to detect preclinical mTBI identified by a battery of cognitive tests. A comparison is performed of the effectiveness of vocal tract dynamics features versus articulatory coordination features. This evaluation is done using receiver operating characteristic (ROC) curves along with confidence bounds. The articulatory dynamics features achieve area under the ROC curve (AUC) values between 0.72 and 0.98, whereas the articulatory coordination features achieve AUC values between 0.94 and 0.97.

Index Terms: concussion, formant dynamics, formant coordination, speech analysis, mild traumatic brain injury (mTBI)

1. Introduction

Traumatic brain injury is known to affect various components of voice including the source, prosody, and articulation of production. These observations have been exploited in speech pathology for diagnosis [1] and in automatic detection of traumatic events from vocal features [2]. In this paper, we focus on the articulation component of production, as manifested by vocal tract formant frequencies. Previous studies have shown that changes in average formant-frequency dynamics [3] or in a novel characterization of articulatory coordination using formant-track correlations [4] can be used to identify or predict depression severity as determined by standard clinical assessment. This paper investigates both of these vocal tract formant representations in a population showing non-physician assessed cognitive impairment. This is a common occurrence among football and soccer athletes, who regularly receive head impacts [5][6]. While each impact may not cause a concussion, also known as a mild traumatic brain injury (mTBI), repeated exposure to head impacts can cause neurocognitive and neurophysiological impairments prior to concussion (preclinical mTBI) [6]. In order to limit the long-term neurological deficits due to repeated preclinical injury, we aim to develop an easily obtainable biomarker for detecting cognitive change. Specifically, we hypothesize that a classifier trained using formant track dynamics and coordination will be able to reliably predict cognitive decline. Such a classifier could detect preclinical damage that accumulates prior to concussion, and determine readiness to return to activity by identifying the point at which an individual has returned to baseline neurocognitive function.

This work is motivated by prior clinical and automatic classification research, which has demonstrated a relation between head trauma and signals associated with vocal dynamics [1][2]. Using speech collected from a group of 32 high school athletes reciting a standard read passage, we extract formant tracks and measure dynamics- and coordination-based features. These features are applied to a support vector machine (SVM) classifier to detect changes in cognitive performance, determined by the Immediate Post-Concussion Assessment and Cognitive Testing (ImpACT) suite [7] in three areas: visual memory, visual motor speed, and reaction time. The ability to detect cognitive performance decline based on these features is evaluated using receiver operator characteristic (ROC) curves.

The paper is organized as follows. In Section 2, we describe the high school football- and soccer-athlete data collection, including the ImpACT cognitive test. Section 3 provides methods to extract features based on dynamics and coordination of formant tracks, while Section 4 gives results of correlating these features with components of the ImpACT test. Guided by correlation results, Section 5 describes a two-class (presence or absence of cognitive impairment) classifier design, and its performance. Finally, Section 6 relates our approach to prior work, with Section 7 describing future plans.
2. Data collection

2.1. Protocol

The data for this study was collected monthly at Purdue University under an institutional review board (IRB) approved protocol. The study includes pre-season, in-season, and post-season data from 32 high school athletes, of whom 25 are male football players and 7 are female soccer players. The athletes’ ages range from 15 to 18, with all data collected independently of any clinical diagnoses of concussions. For each athlete, the data collection includes scores from the online ImPACT assessment version 2.1, which comprise a series of cognitive tests typically used in the sports community [7] (described below), along with voice recordings, eye-tracking, auditory perception, and optic nerve sheath diameter measurements, all of which are part of a multi-modal collection suite developed at MIT Lincoln Laboratory.

This collection platform, referred to as the Multimodal Early Detection Interactive Classifier (MEDIC) system, is designed to extract features from multiple modalities relevant for detecting mTBI [8]. This system provides features that assess voice, aural perception, optic nerve sheath diameter, and ocular dynamics, with the vocal modality being the focus of the current study. Speech features were extracted from audio recordings of the Grandfather Passage, which provides a standardized and phonetically balanced sample of speech.

2.2. ImPACT cognitive assessment

The Immediate Post-Concussion Assessment and Cognitive Testing suite (ImPACT) was used as a means of assessing cognitive performance. The test is made up of six sections, which measure verbal memory, visual memory, visual motor speed, reaction time, impulse control, and a total symptom composite. Because previous studies [9] show that the greatest consistency and correlation to mTBI is with the visual memory, visual motor, and reaction time composites, our study focuses on these test sections. The visual memory section evaluates attention, scanning, learning, and memory. The visual motor speed section evaluates visual processing, learning, memory, and visual motor speed. The reaction time composite is designed to evaluate the subject’s average response time.

3. Feature extraction

3.1. Formant dynamics

The vocal features are based on formant tracks, which are extracted based on the principle that formants are correlated with one another in both frequency and time [10]. Figure 1 shows an example of formant tracks, where the shaded areas around each track represent the 3-dB formant bandwidth. Embedded in the algorithm is a speech-activity detector that enables a Kalman predictor to smoothly coast through non-speech regions.

To characterize formant dynamics, we extract nine formant functions over 20-ms segments at 10-ms frame intervals. These functions are the three raw formant tracks and their high-pass and low-pass components. The common 3-dB cutoff frequency of the high- and low-pass filters is 55 Hz. From each of the above functions we compute three dynamics functions: the raw function value, the velocity, and the acceleration. Derivatives are computed by averaging the first forward and backward differences across frames. The set of features characterize overall formant dynamics, their low-frequency target trajectories, as well as their variability around these targets. The high-frequency component captures the transitional formant properties that are more high-pass in nature, whereas the low-frequency component captures formant properties that are perhaps more semantic-based, representing a smoothed rendition of underlying planned targets. Statistical features used to summarize the dynamics features are: mean, variance, and energy. The resulting number of features is 81: 3 statistical features × 3 dynamics functions × 3 frequency functions × 3 formant tracks.

![Figure 1: Example spectrogram of the first three formant tracks and cross-correlation calculations. Shaded blue region around each track depicts 3-dB formant bandwidth. Channel-delay correlation matrix is the output in the lower right (also see Figure 3).](image)

3.2. Formant coordination

Formant frequency correlations are also investigated under the hypothesis that mTBI may degrade the physiological coordination underlying vocal tract trajectories. We use a multivariate feature extraction approach to assess this degradation, in which changes in coordination are manifested in changes to correlations among the formant tracks. This feature extraction approach has previously been applied to the prediction of major depressive disorder from vocal signals, and was a key component of the winning submission in the Audio Visual Emotion Challenge (AVEC) 2013 depression recognition sub-challenge [4]. A detailed description of this feature analysis approach, in the context of epileptic seizure prediction from multichannel EEG, is provided in [11].

In this approach, channel-delay correlation and covariance matrices are computed from the first three formant tracks (see Figure 1 for schematic). Each matrix contains correlation or covariance coefficients between the tracks at multiple relative time delays. The approach is motivated by the observation that auto- and cross-correlations of measured signals can reveal hidden parameters in the stochastic-dynamical systems that generate the signals. Changes over time in the coupling strengths among the formant tracks, which are sampled at 10-ms frame intervals, cause changes in the eigenvalue spectra of these channel-delay matrices. The matrices are computed at four separate sub-frame “scales,” with 15 time-delays used per scale. The four scales correspond to sub-frames that are spaced 1, 3, 7, and 15 frames apart (i.e., 10, 30, 70, and 150ms frame intervals). Two additional summary statistics are also extracted at each scale from the channel-delay covariance matrices, which characterize the overall covariance power and entropy computed as the log of the trace and determinant of...
the covariance matrix, respectively. The resulting number of features is 188: 4 scales × (15 delays/scale × 3 formants + 2). In Section 5, we discuss an approach to reduce this high dimensionality.

4. Feature analysis

4.1. Correlation analysis of formant dynamics features

The correlations of the formant frequency dynamics features were calculated against ImPACT. Figure 2 shows an illustration of correlations using the Grandfather passage (described in Section 2.1) and features based on the first vocal tract resonance: Variance of the formant frequency, its velocity, its high pass component, and the velocity of its high pass component. One interpretation of these results is that as reaction time decreases articulatory dynamics increase; likewise, as visual acuity increases articulatory dynamics increase.

![Figure 2: Correlations of the formant features extracted from the grandfather passage with sections of ImPACT assessment.](image)

4.2. Analysis of articulatory coordination

Examples of the formant coordination features are given in Figures 3 and 4 with analysis over the Grandfather passage.

![Figure 3: Channel-delay correlation matrix for 70-ms frame interval (3rd scale), computed from formant tracks, associated with cognitive decline (left) and normal function (right).](image)

Figure 3 displays two channel-delay correlation matrices for the 70-ms frame interval (3rd scale). The matrices each contain nine 15×15 blocks, with each block consisting of the within- or cross-channel correlation coefficients for one pair of formant tracks. These coefficients are computed using all possible pair-wise combinations of the 15 time-delayed versions of each formant track. The 15×15 blocks along the main diagonal contain the within-formant correlations and the 15×15 off-diagonal blocks contain the cross-formant correlations. The matrix on the left is from a test session where the subject showed a preclinical effect, while the matrix on the right is from a normal session. The matrix from the normal session appears to contain auto- and cross-correlation patterns with greater complexity, in the sense of larger energy in high-frequency components.

![Figure 4: Eigenspectra example (left) and their average z-normalized renditions (right) from formant channel-delay matrices associated with cognitive decline (red) and normal function (blue).](image)

The qualitative differences in the appearances of the two matrices can be quantified using the matrix eigenspectra, which are the rank-ordered eigenvalues. The eigenspectra from the two matrices are shown in Figure 4 (left), with the eigenvalues associated with decreased performance in red and normal performance in blue. The eigenspectra from the impaired subject contain a smaller fraction of power in the higher eigenvalues, indicating reduced complexity and independent variation in this subject’s formant tracks. This example is consistent with the average pattern across all subjects, shown in Figure 4 (right), in which the z-normalized eigenvalues are plotted in standard units.

The divergence in eigenspectra between the high-and low-score subjects is qualitatively similar to the divergence that has previously been found between normal and depressed subjects [4]. This commonality suggests that the present measures of formant dynamics and coordination may constitute a robust indicator of the health of neurological motor systems.

5. Classifying change in cognitive performance

Features extracted from voice are used to classify changes in cognitive performance. Cognitive performance is measured using three sections of the ImPACT test, which have been shown to have high consistency and validity in previous studies [7]. As discussed in Section 2.2, these include the visual memory, visual motor speed, and reaction time components. Using each test, a threshold is set for a change in cognitive performance. The threshold for each test is defined as a decline from baseline that exceeds one standard deviation, where the standard deviation is computed over the change from baseline across all subjects’ test scores. Additionally, we predict changes for a combined set, which labels a change in performance if there is a suprathreshold decline in any of the three tests described above. The above procedure results in 13 changes for visual memory, 5 for visual motor speed, 9 for reaction time, and 23 for the combined set, out of a total of 78 subject recordings. A support vector machine (SVM)-based classifier is then used to predict, based on our formant features, a decline in cognitive performance.
5.1. SVM classifier

This study’s prediction model is implemented using a two-class support vector machine (SVM) classifier. The SVM was implemented using the model developed by Chang et al. [12].

Vocal features were first z-normalized before principal component analysis (PCA) was applied. PCA was applied separately to the formant track dynamics and incoordination features. In order to limit the number of features and prevent over-fitting, a subset of the principal components was used to train the classifier. The number of principle components was chosen empirically with a constant number of components used for the three cognitive tests, and a separate number of components for the combined set. From the dynamics features, seven components were used for the individual tests and five components for the combined test. From the coordination features, nine components were used for the individual tests and seven components for the combined test. The PCA features were once again z-normalized before training the classifier, with a radial basis function used to kernelize the feature space. The parameters for this transformation were selected using a cross validation loop.

Training and testing for the SVM are done using cross-validation, with a single subject being held out for testing at each iteration. A classifier is formed for each cognitive assessment.

5.2. Results

ROC curves (Figure 5), along with the true positive and false alarm rates, and the area under the curve (AUC), are used to evaluate classification performance. True positives are when the classifier correctly predicts a decline from baseline in an athlete’s cognitive performance, whereas false alarms are when it incorrectly predicts a decline. Additionally, a bootstrapping procedure with 1000 intervals is used to calculate lower and upper bounds for the AUC.

<table>
<thead>
<tr>
<th>Features</th>
<th>ImPACT</th>
<th>AUC</th>
<th>AUClow</th>
<th>AUChigh</th>
<th>FA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dynamics</td>
<td>VisMem</td>
<td>0.95</td>
<td>0.87</td>
<td>0.98</td>
<td>0.15</td>
</tr>
<tr>
<td>Coordination</td>
<td>VisMem</td>
<td>0.94</td>
<td>0.87</td>
<td>0.98</td>
<td>0.20</td>
</tr>
<tr>
<td>Dynamics</td>
<td>VisMotor</td>
<td>0.96</td>
<td>0.94</td>
<td>1.00</td>
<td>0.03</td>
</tr>
<tr>
<td>Coordination</td>
<td>VisMotor</td>
<td>0.95</td>
<td>0.87</td>
<td>0.98</td>
<td>0.07</td>
</tr>
<tr>
<td>Dynamics</td>
<td>Reaction</td>
<td>0.77</td>
<td>0.64</td>
<td>0.88</td>
<td>0.39</td>
</tr>
<tr>
<td>Coordination</td>
<td>Reaction</td>
<td>0.97</td>
<td>0.90</td>
<td>0.99</td>
<td>0.10</td>
</tr>
<tr>
<td>Dynamics</td>
<td>Union</td>
<td>0.72</td>
<td>0.57</td>
<td>0.87</td>
<td>0.56</td>
</tr>
<tr>
<td>Coordination</td>
<td>Union</td>
<td>0.95</td>
<td>0.88</td>
<td>0.98</td>
<td>0.27</td>
</tr>
</tbody>
</table>

Table 1: Classification performance using AUC and false-alarm rate to get 1.00 detection rate (FA).

Table 1 displays the AUC values, along with its lower and upper bounds. The minimum false alarm rate to obtain a 1.00 true positive rate is also displayed. As there is high incentive to identify cognitive decline and prevent long-term neurological impairment, this number reflects the percent of false alarms incurred to ensure perfect detection. Comparing across ImPACT components, the dynamics features show near identical performance to the coordination features for the visual memory and the visual motor tasks. However, the coordination features provide a strong improvement over the dynamics features on both the reaction time component of ImPACT and on the combined set. On the combined set, the simple statistics have a lower bound performance close to chance, while the coordination features’ lower bound is at 0.88. The coordination features therefore provide a robust basis for classifying cognitive change across tests.

6. Discussion and relation to prior work

mTBI is a serious problem for athletes as well as others exposed to head impacts. Recent research has focused on gaining a better understanding of the effects of mTBI, as well as on providing additional utilities to diagnose and prevent subsequent injury. Cao et al. [13] used a SVM classifier along with EEG to identify functional abnormalities in the brain. Falcone et al. [2] used a classifier to identify when a concussion has occurred using the speech produced during held vowels. Our current study is in contrast to this work through its use of vocal features during read speech to detect changes in cognitive ability. Detecting changes in cognitive status has potential benefit, in that cognitive changes have been shown to arise prior to clinically diagnosed concussions [6], which could lead to an increased susceptibility to mTBIs. Furthermore, the high detection rate of the classifier suggests it could be used as a screening tool to determine readiness to return to activity, thereby lowering the subject’s risk for subsequent injury.

7. Future work

In future work, we will exploit a suite of features from source, prosodic, and articulatory aspects of voice, as well as eye-tracking features and fMRI data. Correlation of features with brain imaging data will help improve understanding of the effects of head impacts on functional activations.

8. Acknowledgements

The authors thank Dr. Jonathan Su, Adam Dai, Daniel Strom, Andreas Gennis, Dr. Daryush Mehta, and Dr. Nicolas Malyska for their involvement and contribution to this work.
9. References


