Noninvasive Biomarkers of Neurobehavioral Performance

Thomas F. Quatieri, James R. Williamson, Adam C. Lammert, Kristin J. Heaton, and Jeffrey S. Palmer

A large array of neurological and psychological conditions is prevalent in civilian and military populations. To complement current clinical standards, there is a pressing need for noninvasive in-field and at-home methods of assessing such conditions. Lincoln Laboratory is developing neurobehavior-based biomarkers, which reflect a change in brain functioning as manifested in motor control, paired with neurocomputational biophysical models to identify neurobehavioral changes. The efficacy of our approach is illustrated through several applications for assessing major depressive disorder, Parkinson's disease, traumatic brain injury, and cognitive overload.

» **Mental health impairments can** significantly and adversely affect human performance and quality of life for civilian and military populations. Indeed, brain disorders taken together are a leading cause of global disease [1]. Toward the goal of finding simple, noninvasive, and objective means to detect, predict, and/or monitor such conditions, the Human Health and Performance Systems Group at Lincoln Laboratory is developing multimodal biomarkers based on behavioral measurements to detect changes in neurobehavioral function associated with psychological (e.g., major depressive disorder), neurotraumatic (e.g., traumatic brain injury), neurodegenerative (e.g., Parkinson's disease), and neurodevelopmental (e.g., autism spectrum disorder) conditions.

One of our objectives is the discovery of behavioral-based biomarkers that reflect a change or decline in brain functioning as manifested in motor control and, more specifically, changes in timing and coordination within the neuromotor components of a behavior. This investigation into biomarkers is based on the hypothesis that motor control is associated with neural coordination across different parts of the brain. Examples of behaviors we are examining include vocal and facial expression, heart rate variability, eye movement, and fine and gross movement of the extremities. One focus of our investigation is on novel vocal and facial biomarkers that are based on phonetic timing, articulatory coordination, and facial muscle coordination during speech production.

A motivation for working in the areas of health and performance is the prevalence in civilian and military

populations of conditions that can compromise neurobehavioral function. Among the neurobehavioral disorders that affect global health, depression is the single largest source of lost productivity in high-income countries. According to the World Health Organization [1], by the year 2030, disability and lives lost from depression alone will surpass those caused by cancer, war, stroke, and accidents. There are roughly 20 suicides per day among U.S. veterans [2]. It also is projected that by 2030 there will be 82 million people worldwide with dementia [3]. One in 59 children in the United States has been identified as having autism spectrum disorder (ASD) [4]. About 19 percent of U.S. servicepeople returning from recent wars experienced a traumatic brain injury (TBI) of some form, and about 20 percent of that same population has post-traumatic stress disorder (PTSD) [5]. Other factors, such as (1) exposures to environmental extremes or occupational chemicals and (2) fatigue resulting from physical and mental exertion or disrupted sleep patterns, can also compromise neurobehavioral performance in our servicepeople. Figure 1 illustrates the staggering and increasing prevalence of mental health conditions and other factors that can impair neurobehavioral performance.

The many standard methods used in detecting neurobehavioral performance changes range from brain imaging to clinical assessments to molecular diagnostics. These approaches, while useful, can be time and resource intensive, are often susceptible to the effects of individual motivation, and, because they often capture feelings or behaviors at discrete points in time, may lack sufficient sensitivity to detect subtle changes in behavior. Moreover, they often lack objective measurement, especially in early detection of day-to-day performance changes when assessment can be most important. At Lincoln Laboratory, we are developing technologies that address these limitations by reaching large populations and detecting, monitoring, and ultimately intervening to follow the effect of treatment and intervention. Our approach seeks objective biomarkers that reflect subtle changes in behavior and makes use of nonobtrusive mobile wearable technologies.

In the context of this article, the term biomarker refers to any measurement of behavior we can obtain from a human body, such as talking or walking. Certain behavioral features have been shown to change with a subject's mental and emotional state and under numerous conditions, including cognitive load and neurological

FIGURE 1. An array of neurological and psychological conditions is prevalent in U.S. civilian and military populations.

disorders. Example modalities used in detecting cognitive and neurological stress include voice, facial expression, physiology, eye movement, gait, hand dexterity, and electroencephalography (EEG) analysis. The use of behavioral-based biomarkers is not a new concept. Such biomarkers have been applied in a variety of contexts for at least two decades. Examples include features derived from vocal and facial expression, fine and gross motor movements, physiology, and brain activity [6–40].

Features derived from vocal expression, or voice, include characterizations of prosody (e.g., fundamental frequency and speaking rate), spectral representations (e.g., vocal tract resonance), and glottal excitation flow patterns (flow shape, timing jitter, amplitude shimmer, and aspiration) [10, 13, 14, 33, 36] derived from acoustic measurements. Features from facial expression include parametric models of the face and representations of underlying facial muscle groups known as facial action units [22, 24] derived from video measurements. For fine and gross motor movements, irregularities in stride, hand dexterity, and eye tracking have been used in characterizing a variety of neurological conditions, with a few examples in Russo et al., Samadani et al., and Gowan and Hamilton [40–42].

We cannot hope in the limits of this article to review all state-of-the-art behavioral-based feature approaches. However, what distinguishes Lincoln Laboratory's approach from these more standard methods, or what we will refer to as low-level feature-based approaches, is the introduction of features motivated by the timing and coordination of underlying neuromotor control of behaviors. Although significant effort has been devoted to behavioral-based biomarkers, little or no study has been done examining changes in coordination, movement, and timing of components of a behavior. For example, in individuals suffering from depression, neurophysiological changes often alter motor control and thus affect mechanisms controlling speech production and facial expression. Clinically, these changes are typically associated with psychomotor slowing, a condition of slowed neuromotor output manifested in altered timing and coordination across multiple observables of acoustics and facial movements during speech. We refer to features based on this paradigm as high-level features.

While we begin in this article with standard low-level features in each modality, we build upon these using high-level timing and coordination features that reflect underlying neural activity across the brain. It is hypothesized that these relations are associated with neural coordination across different parts of the brain that are essential in motor control. Various subsets of these features have been used effectively at Lincoln Laboratory for detecting neurobehavioral changes associated with depression, Parkinson's disease, TBI, and dementia [13, 43–46], as well as mental exertion under stress [43, 47], and thus perhaps form a common feature basis for neurocognitive change.

Neurophysiological Basis and Framework

Our approach involves biomarkers of human behavior that we can observe from the human body. At the Laboratory, we are investigating a number of different behaviors, broadly defined to include fine motor movements (e.g., hand and finger dexterity and eye tracking), gross motor movements (e.g., balance and gait), skin conductance, and heart rate. But we are focusing in this article on vocal and facial expression, and specifically on markers that reflect change in motor aspects of brain function.

Neuromotor Representations in the Brain

In our approach, we seek biomarkers that satisfy two primary properties:

1. They reflect decline in brain functioning as manifested in motor control measured from bodily behaviors.
2. They reflect changes in timing and coordination both within and across components of a behavior.

Deriving biomarkers from vocal and facial expression is desirable for a number of reasons: vocal and facial expressions are easily measured, noninvasive, and accessible, and, importantly for our approach, they are highly complex human behaviors that require precise coordination across different regions of the brain. In speaking, for example, the articulators (tongue, lips, jaw, and velum) are finely coordinated, and this coordination can change under conditions of injury, illness, and stress. Underlying this articulatory complexity is the even more complex control system of the brain [48]. As an example, we overview this control system in speech production. There is evidence of a similar kind of neural network complexity that controls facial expression during speaking, as well as in general expressiveness in paralinguistic socioemotional communication [49].

As schematized in Figure 2, the core speech-production network consists of many components, widely distributed throughout the brain. These components go from developing a concept and selecting sentences and words to express a concept, to deciding syllables and phonemes to represent words, to positioning and coordinating articulators, to firing neural signals that activate muscles to move the articulators. Finally, there are auditory and somatosensory feedback mechanisms used to monitor and self-correct speech production. Somatosensory includes proprioception (a person's sense of position and movement of the body) and tactile feedback (a person's sense of tongue placement and vocal cord vibration).

In addition to this core production network, nonspeech regions of the brain modulate the core. Given this wide distribution of modules in the brain that either directly or indirectly control speech production, it is likely that one of the modules will be affected by neurological or stress conditions, and speech production will suffer. In fact, all of the brain disorders mentioned at the start of this article cause some kind of speech degradation. For depression, as an example, there is evidence that a nonspeech region called the limbic system, which controls mood and emotion, is disrupted. This disruption propagates to the core speech network, in particular to neural circuits that control timing and coordination of the articulators. The general observation by clinicians that depressed individuals, on average, tend to talk slower and have less clarity in their articulation than do nondepressed individuals is consistent with a disruption of timing and coordination circuits. Motivated by these observations, we have introduced two novel biomarkers that are based on the decline in neuromotor timing and

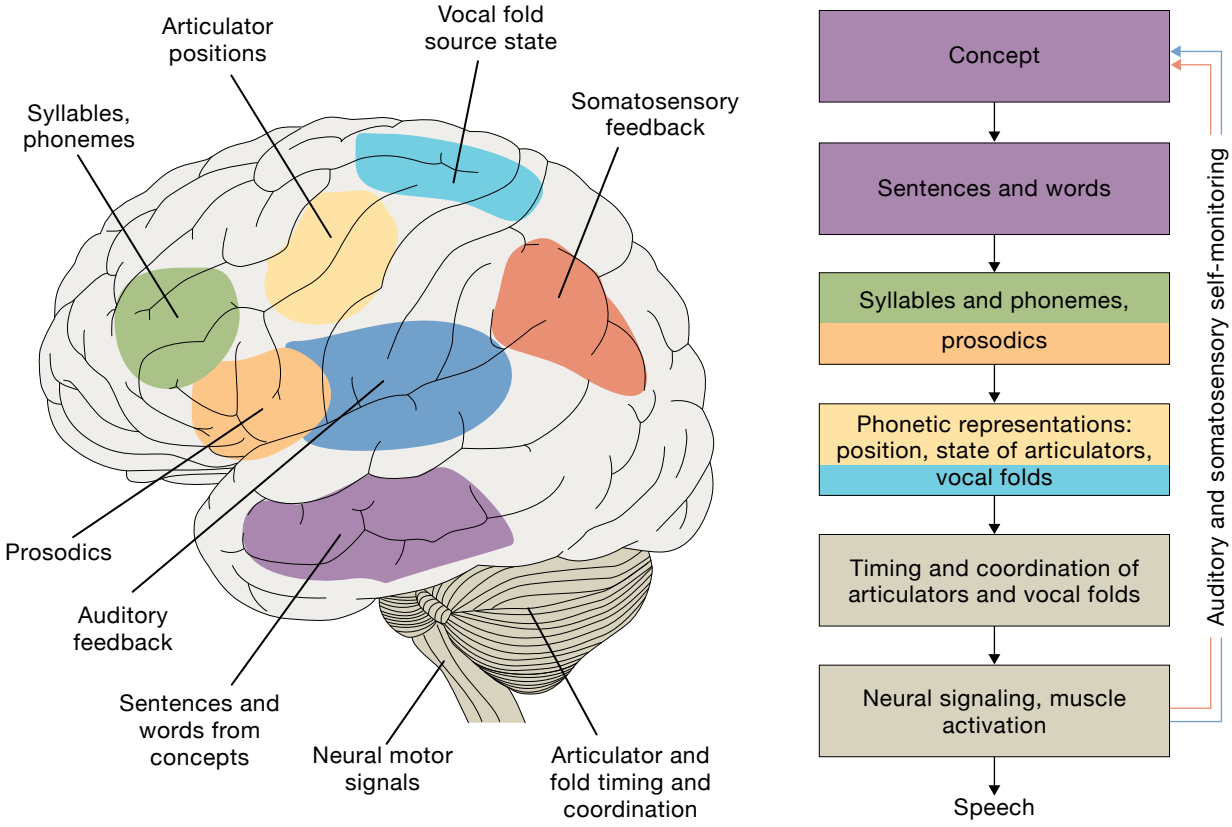


FIGURE 2. The figure shows a simplified view of the core speech production neurocognitive network with many components widely distributed throughout the brain. The somatosensory region is associated with tactile and proprioceptive (a sense of position and movement of the articulators) sensory feedback. Nonspeech regions, such as visual and cognitive areas (not shown), can modulate the core speech network. It is important to note that Figure 2 is a gross oversimplification of the speech brain network that involves complex coordination across multiple brain regions.

coordination that we will describe. We also consider these biomarkers more generally across all conditions.

Framework

The new features and the corresponding detection system that we have designed leverage many years of Laboratory expertise in speech and facial signal processing and automatic classification. At a bird's eye view (Figure 3), our detection system first estimates a set of standard vocal and facial features. These features that leverage our and others' past work [50–52] we call low-level features. Examples of low-level features from an acoustic speech signal are vocal tract resonances (termed formants in the speech community) and automatically derived speech phoneme labels (units of sound that distinguish one word from another). Low-level features from facial video are facial action units that reflect muscle groups associated with basic elements of expression (lowered brow, puckered lips). From these low-level features, we then extract our new features on the basis of a timing and coordination concept, and we call these high-level features. With the high-level features, we train a classifier that provides us with a binary decision or severity-level estimate of a condition.

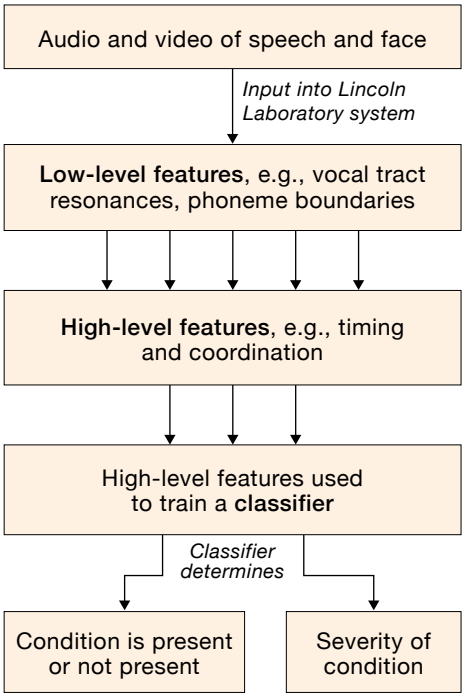


FIGURE 3. Lincoln Laboratory's system can detect the presence or estimate the severity of a neurobehavioral condition. Low-level features are first derived from audio and video of speech and face, respectively. Next, high-level features are extracted from neuromotor timing and coordination features. These high-level features are then used to train a classifier to determine the presence and/or severity of a neurobehavioral condition.

Feature Extraction

Vocal Features

The categorization of vocal characteristics is broken down into three components: speech source (at the vocal folds), system (vocal tract), and prosody (sometimes referred to as the melody of speech). As shown in Figure 4, the lungs provide the airflow that assists in making the vocal folds vibrate, sending a periodic or noise-like stimulus to the vocal tract. The vocal tract—consisting of the oral, nasal, and pharynx subsystems—provides "color" to the sound, with different tract shapes yielding different phonemes. Prosody is a function of pitch, timing, and energy fluctuations.

We exploit dynamic variation and interrelationships across speech production systems by computing features that reflect complementary aspects of the speech vocal-fold source, vocal tract system, and prosody [50]. Across the three categories, we use a broad suite of features that are used in the detection of neurological disorders and neurobehavioral performance changes associated with a wide range of exposures and other sources of stress [13, 43–47, 53–60]. In each category, we leverage low-level and high-level feature types that were introduced previously.

LOW-LEVEL FEATURE EXTRACTION

The various low-level features are estimated from windowed speech segments frame-by-frame over time, and the window type (e.g., Hamming, Hanning, Kaiser), length (typically between 10–40 ms), and sliding frame (typically between 1–10 ms) are selected depending on the feature type.

Many standard low-level features used in the speech community characterize the degree of periodicity of the vocal-fold vibration within the vocal source. One such feature is the harmonics-to-noise (HNR) ratio, which is the ratio, in decibels (dB), of the power of the harmonic (periodic) signal from vocal-fold vibration and the power of the speech noise signal at the vocal folds created by turbulence as air rushes past the vocal

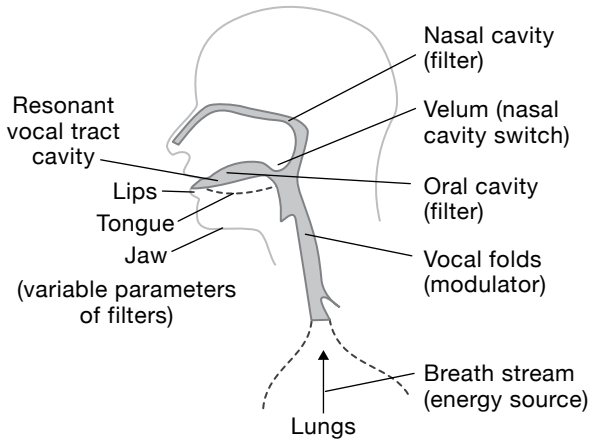


FIGURE 4. This illustration shows the speech source (at the vocal folds) and system (vocal tract) [50]. Prosody is a function of pitch, timing, and energy fluctuations. The lungs provide an energy source and the airflow that assists in making the vocal folds vibrate, providing input to the vocal tract. The vocal tract—consisting of the oral, nasal, and pharynx subsystems—provides "color" to the sound created at the vocal folds, with different tract shapes yielding different phonemes.

folds from the lungs. This measure is thought to reflect breathiness in a voice. Our HNR is computed by using a periodic/noise decomposition method that employs a comb filter to extract the harmonic component of a signal [61, 62]. Two other low-level source features are (1) cepstral peak prominence (CPP), which is defined as the difference in dB between the magnitude of the highest peak and the noise floor in the power cepstrum (the result of taking the inverse Fourier transform of the logarithm of the estimated spectrum of a signal) and (2) a measure called creak, corresponding to what is often referred to as creaky voice, which reflects large irregularity in pitch periods (often with low average pitch) and high peakiness of airflow pulses that excite the vocal tract [63, 64].

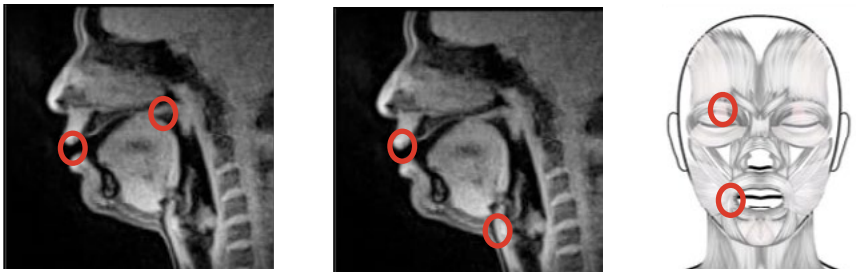
Low-level vocal tract–based features are designed to reflect the intensity and temporal dynamics of the vocal tract frequency response. One primary feature set comprises the vocal tract resonances (referred to as formant frequencies) estimated by a Kalman filter technique, smoothly tracking the first three formant frequencies while also smoothly coasting through nonspeech regions [28]. A second primary feature set is referred to as Mel-frequency cepstral coefficients (MFCCs) that provide frequency response intensity, while 16 delta MFCCs [50] are used to characterize velocities of vocal tract spectral magnitudes [65]. Delta MFCCs are computed by using regression with the two frames before and after a given frame.

The final low-level feature types reflect speech prosody. The first feature type is pitch (also referred to as fundamental frequency), which we estimate by using a time-domain autocorrelation method [50]. Our second prosodic-based feature type relies on an automatic phoneme recognition algorithm [66]. This algorithm obtains phonetic boundaries and phoneme labels with one of 40 phonetic speech classes detected and allows us to obtain average speaking rate (number of phonemes per second).

HIGH-LEVEL FEATURE EXTRACTION

Our high-level features, often derived from low-level features, are designed to capture timing, coordination, and fine time resolution of dynamics of speech production components. We refer to one high-level feature type as correlation structure, which is a function of the temporal correlation (on different time scales), reflecting a form of coordination within and across vocal source, system, and prosodic speech components (illustrated in Figure 5).

In this approach, channel-delay correlation and covariance matrices are computed from multiple time-series channels of vocal parameters. 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 multiple time scales corresponding to separate subframe spacings. Features at each time scale consist of the eigenvalue spectra of channel-delay correlation matrices, covariance power (logarithm of the trace), and entropy (logarithm of the determinant) from channel-delay covariance matrices. Under various conditions, we find different degrees of dynamical complexity both within (e.g., formant frequencies, or HNR versus CPP) and across (e.g., formant frequencies vs creak), dependent on the condition, as reflected in eigenvalue distributions. This methodology is illustrated in Figure 6 with the example of the generation of formant track correlation matrices. Further mathematical details of this approach are in Williamson, Bliss, et al. [58],



(a) (b) (c)

FIGURE 5. Circled in red, the anatomical regions where coordination and timing features are derived are within the articulatory elements that form the vocal tract (a), across vocal tract and vocal source, or fold, components (b), and across muscle groups in facial expression (c). MRI IMAGES COURTESY OF UNIVERSITY OF SOUTHERN CALIFORNIA.

which introduced the method for the analysis of EEG signals for epileptic seizure prediction, and in Quatieri, Williamson, et al. [43], which describes the approach in the context of speech and face processing.

We have also introduced a feature set that characterizes the structure of signal coherence and power at multiple frequency bands within and across speech components over time. We refer to this as coherence structure. The coherence between channels, indicating the amount of cross-channel power in a frequency band relative to the amount of within-channel power, provides a measure of how closely related the signals are within a frequency band. The power and cross-power are computed among three formant frequency channels in three different frequency bands, and a 3×3 coherence matrix is constructed for each band. Analogous to correlation structure, the eigenspectra of the coherence matrices indicate the structure of coherence across the channels.

Our high-level vocal source and prosodic features rely on their low-level counterparts described earlier. In one case, we leverage a finer time scale on a phoneme level (within phoneme boundaries), using feature sets of phoneme-dependent durations [56]. Based on estimated average durations for each phoneme, the summed average durations of certain phonemes are linearly combined to yield fused phoneme duration measures. A subset of phonemes whose summed durations are highly correlated with an assessment (e.g., known depression or cognitive load level) on a training set is selected to create these fused measures, with weights based on the strength of their individual correlations. This phoneme-dependent feature paradigm is shown in Figure 7. A fused phoneme-dependent pitch slope measure is also obtained by using essentially the same procedure as described above. For each passage in a training set, the set of phonemes with the highest correlating summed pitch slopes is selected.

Facial Features

As with vocal features, high-level facial features rely on low-level features. Likewise, analogous to vocal high-level features, we use correlation-based measures that reflect the coordination of facial muscle groups and a rate measure that reflects the duration of each muscle group.

LOW-LEVEL FEATURE EXTRACTION

Characterizing the effects of neurological disorders on facial movements is an active research area. For example, among people suffering from major depressive disorder, measurable differences have been found in facial expressions [22], including acute reductions in involuntary facial expressions in depressed persons [23] and changes in facial expressions that are imperceptible to clinicians [18]. The facial action coding system (FACS) is a systematic method for quantifying localized components of facial expressions called facial action units (FAUs). Each FAU corresponds to distinct sets of muscle movements of the face [24]. The University of California San Diego has developed a computer expression recognition toolbox (CERT) that provides automatic identification of FAU likelihoods or probabilities from individual video frames [27]. Figure 8 lists the FAUs output by CERT for video-based facial expression analysis.

HIGH-LEVEL FEATURE EXTRACTION

The coordination of facial movements during speech can be measured by using correlation structure features obtained from the multivariate FAU time series described in Figure 8. With the same method we used to assess coordination from audio-based features, we can construct high-dimensional channel-delay correlation matrices in which each matrix element represents the correlation coefficient between two FAU time series (channels) at a particular relative time delay [54]. Analogous to our

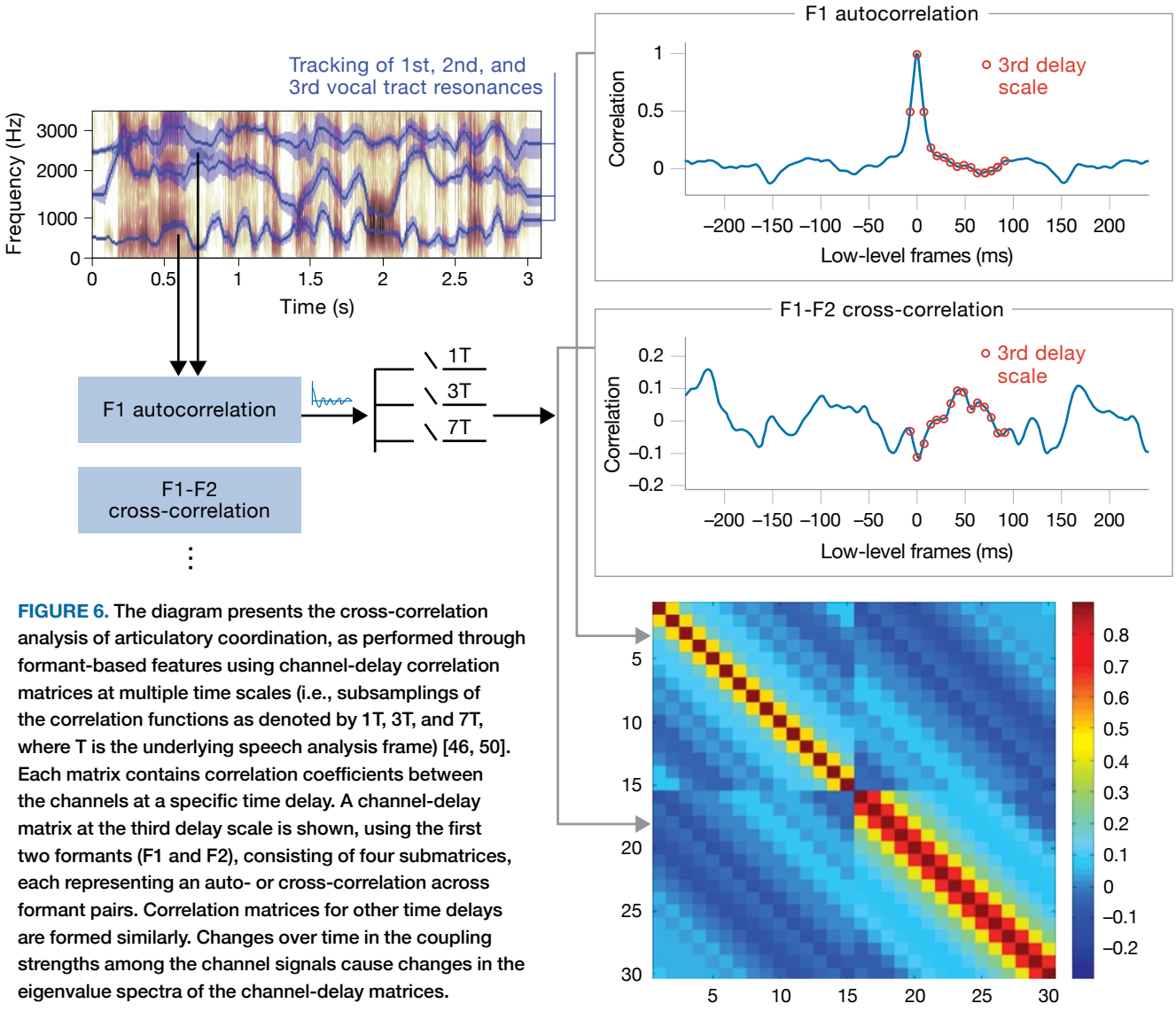


FIGURE 6. The diagram presents the cross-correlation analysis of articulatory coordination, as performed through formant-based features using channel-delay correlation matrices at multiple time scales (i.e., subsamplings of the correlation functions as denoted by 1T, 3T, and 7T, where T is the underlying speech analysis frame) [46, 50]. Each matrix contains correlation coefficients between the channels at a specific time delay. A channel-delay matrix at the third delay scale is shown, using the first two formants (F1 and F2), consisting of four submatrices, each representing an auto- or cross-correlation across formant pairs. Correlation matrices for other time delays are formed similarly. Changes over time in the coupling strengths among the channel signals cause changes in the eigenvalue spectra of the channel-delay matrices.

vocal measures, the eigenspectrum of the correlation matrix then characterizes the total level of independent movement (i.e., complexity) contained in the FAU set.

The facial coordination features represent the total amount of independent movement captured by the FAUs, without regard to the level of movement contained in individual FAUs. Facial activation rate features take a different tack, representing the average level of activation among those FAUs that correlate strongly with the outcome measure of interest, such as depression score [54]. Analogous to our phoneme-duration strategy, an aggregate measure of facial activation rate is then obtained by linearly combining the rates of the highly correlated FAUs, with negative weights assigned to FAUs with negative correlations.

Electroencephalogram Features

Although not often feasible for a long-term objective of a nonintrusive mobile platform, electroencephalogram (EEG) is used as a reference and sometimes as ground truth. As with vocal and facial characterization, we work with low- and high-level features in EEG analysis.

The EEG signals were measured with a 1000 hertz (Hz) sampling with a 64-element Neuroscan system, followed by high-pass filtering and standard artifact removal. The EEG signals were decomposed into five

frequency bands that have been implicated in cognitive, sensorimotor, and perceptual activities (delta, theta, alpha, beta, gamma), with band ranges of 0–4 Hz, 4–8 Hz, 8–16 Hz, 16–32 Hz, and 32–49 Hz, respectively [25].

Decomposition was performed by bandpass filtering in each of these frequency bands at each channel. In our low-level analysis, we use EEG channel-dependent band power to compute measures of spatial activity patterns, which provide a basis for and complementary information to high-level features.

In high-level analysis, our guiding principles are that successful cognition requires coordinated neural activations in brain networks linking multiple brain regions and that these networks communicate using oscillatory codes operating over a wide range of frequencies. Based on these principles, our high-level feature approach is to use measures of neural coordination indicated by EEG connectivity at each frequency band. We use two connectivity measures: pairwise channel coherence and covariance [43, 67]. Coherence measures cross-channel power relative to within-channel power, whereas covariance measures cross-channel correlation weighted by within-channel power.

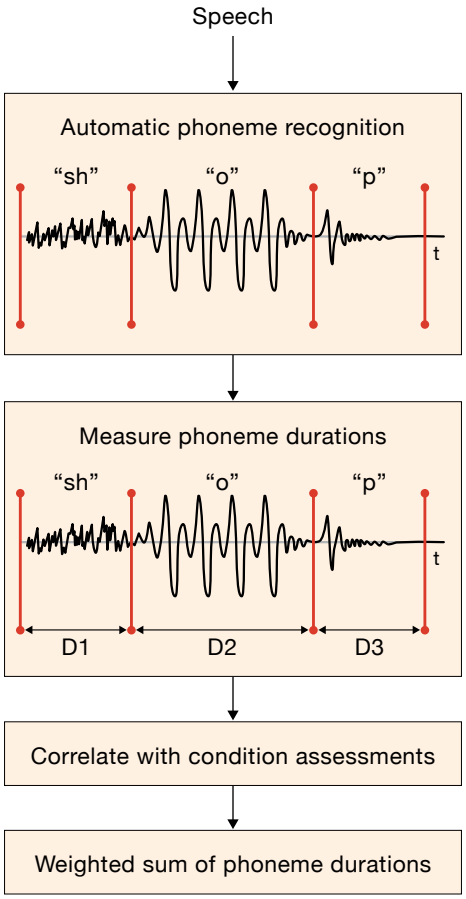


FIGURE 7. Phoneme-dependent duration extraction [56] first requires automatic phoneme recognition, followed by measuring the average duration (D) of each phoneme (there are 42 in the English language). Average duration measures are then correlated with the condition assessment (e.g., severity of depression) across subjects. The final feature is a weighted sum of phoneme durations, where weights are a function of the correlation value for each phoneme.

Neurological Disorders

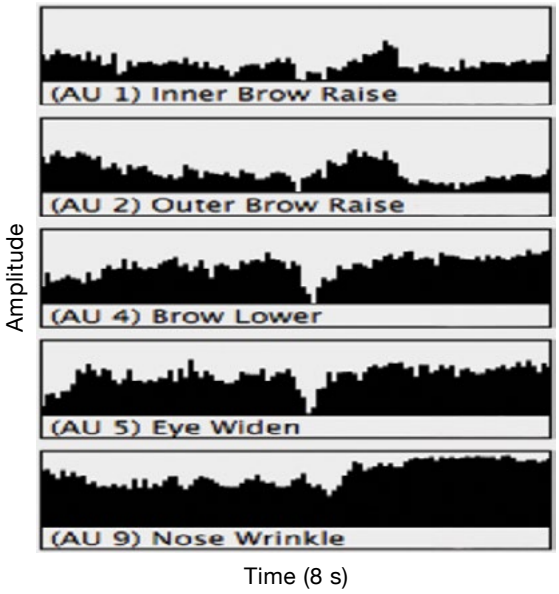
The general classification framework and biomarker extraction approach have been applied to numerous psychological and neurological disorders. Researchers in the Human Health and Performance Systems Group at Lincoln Laboratory have applied these principles to the detection of major depressive disorder, Parkinson's disease, mild traumatic brain injury, dementia, and amyotrophic lateral sclerosis (ALS). Here, we focus on the first three of these conditions.

Predicting Major Depressive Disorder Severity

Major depressive disorder (MDD) is the most prevalent mood disorder, with a lifetime risk of experiencing the disorder ranging from 10 to 20 percent for women and 5 to 12 percent for men [68]. As the number of people suffering from MDD steadily increases, so too does the burden of accurate diagnosis. Currently, the diagnosis of MDD requires a comprehensive assessment by a professional with significant clinical experience. However, the inter-clinician variability of these assessments makes the tracking of medication efficacy during clinical trials difficult. The growing global burden of MDD suggests that a convenient and automated method to evaluate depression severity would both simplify and standardize the tasks of diagnosing and monitoring depression, allowing for greater availability and uniformity in assessment. An automated approach may reduce multiple in-office clinical visits, facilitate accurate measurement and identification, and quicken the evaluation of treatment. Toward these objectives, potential depression biomarkers of growing interest are vocal and facial expression features, two categories of easily acquired measures that have been shown to change with a patient's mental condition and emotional state [14–16, 24, 27, 28, 66, 69].

ACTION UNIT NUMBER	FACIAL ACTION UNIT DESCRIPTION
1	Inner brow raiser
2	Outer brow raiser
4	Brow lowerer
5	Upper lid raiser (eye widen)
6	Cheek raiser
7	Lid tightener
9	Nose wrinkler
10	Upper lip raiser
12	Lip corner puller
14	Dimpler
15	Lip corner depressor
17	Chin raiser
18	Lip pucker
20	Lip stretcher
23	Lip tightener
24	Lip presser
25	Lips part
26	Jaw drop
28	Lip suck
45	Blink/eye closure

(a)



(b)

FIGURE 8. The 20 facial action units from the computer expression recognition toolbox, or CERT, are listed in (a) with their corresponding action unit (AU) number. (We list only the 20 facial action units that are included in CERT; there are many additional facial action units not listed here.) Example time series of five of these facial action units, extracted from video during speaking, are shown in (b). The time series illustrate time variations of each of these facial action units.

VOCAL AND FACIAL CHARACTERIZATION

Depression exhibits changes in all three vocal components described above: speech excitation (source), vocal tract (system), and pattern of stress and intonation (prosody). Depression-related changes in speech reflect the perception of qualities, such as monotony, slur, slowness, hoarseness, and breathiness, in the speech of depressed individuals. Hoarseness and breathiness may be associated with speech source characteristics (at the level of the vocal folds). Monotony may be associated with prosody (e.g., modulation of speech-rate, pitch, and energy), and slur with speech system characteristics (e.g., vocal tract articulators). Likewise, characterizing the effects of depression on a speaker's facial movements is an active research area. Early work found measurable differences between facial expressions of people suffering from MDD and facial expressions of nondepressed individuals [24]. Electromyography monitors can register facial expressions that are imperceptible during clinical assessment [18] and have found acute reductions in involuntary facial expressions in depressed persons [23].

Although significant effort has focused on studying vocal and facial biomarkers for emotion classification, little or no study has been done investigating changes in coordination, movement, and timing by using speech and facial modalities for depression classification or severity prediction. In individuals suffering from MDD, neurophysiological changes often alter motor control and thus affect mechanisms controlling speech production and facial expression. Clinically, these changes are typically associated with psychomotor retardation, a condition of slowed

neuromotor output manifested in altered coordination and timing across multiple observables of acoustics and facial movements during speech [53]. Alluded to earlier is evidence that a nonspeech region called the limbic system (that controls mood and emotion) is disrupted, and this disruption propagates to the core speech network and in particular neural circuits that control timing and coordination of the articulators (Figure 9). Consistent with a disruption of timing and coordination circuits is the general observation by clinicians that depressed subjects, on the average, tend to talk more slowly and have less clarity in their articulation than nondepressed individuals. Likewise, during speaking, the corresponding facial expression is slower and flatter in effect.

FEATURE SELECTION

Motivated by this neural-based hypothesis and clinical observations, we developed and applied a variety of the high-level articulatory and facial coordination features and vocal and facial timing features outlined earlier [46, 53, 54, 56]. Our high-level coordination-based features are designed to characterize properties of coordination from the low-level features. After investigating multiple combinations of the low-level vocal features as input to correlation analysis, we found the best overall performance was achieved by using the following three combinations: (1) formant–CPP, (2) CPP–HNR, and (3) delta MFCC. Channel-delay correlation and covariance matrices are computed from multiple time-series channels (of given vocal and facial parameters). Each matrix contains correlation or covariance coefficients between the channels at multiple relative 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 frame spacings of 1, 3, 7, and 15. Overall covariance power (logarithm of the trace) and entropy (logarithm of the determinant) are also extracted from the channel-delay covariance matrices at each scale. For vocal-based timing features, we use cumulative phoneme-dependent durations and generalize to phoneme-dependent pitch slopes, obtained by using estimated phoneme boundaries. For facial-based timing features, we use FAU rates obtained from their estimated posterior probabilities.

EXAMPLE DETECTION

We have tested our biomarkers on a variety of depression databases. Here we focus on the 2014 Audio/Video Emotion Challenge (AVEC) that uses a depression corpus that includes audio and video recordings of subjects performing two human-computer interaction tasks in the German language: (1) reading a phonetically balanced passage and (2) replying to free-response questions [54]. Data were collected from 84 German subjects with ages ranging between 18 and 63 years, with a mean of 31.5 years and a standard deviation of 12.3 years. Video of the subjects' faces was captured with a webcam at 30 frames per second and a spatial resolution of 640 × 480 pixels. Audio was recorded with a headset microphone at a sampling rate of 32 kHz or 48 kHz. For each session, the standard self-reported Beck Depression Inventory (BDI) assessment score was available. The recorded sessions were split into two datasets: 100 subjects in the training set for designing a classifier and 50 subjects for the test set.

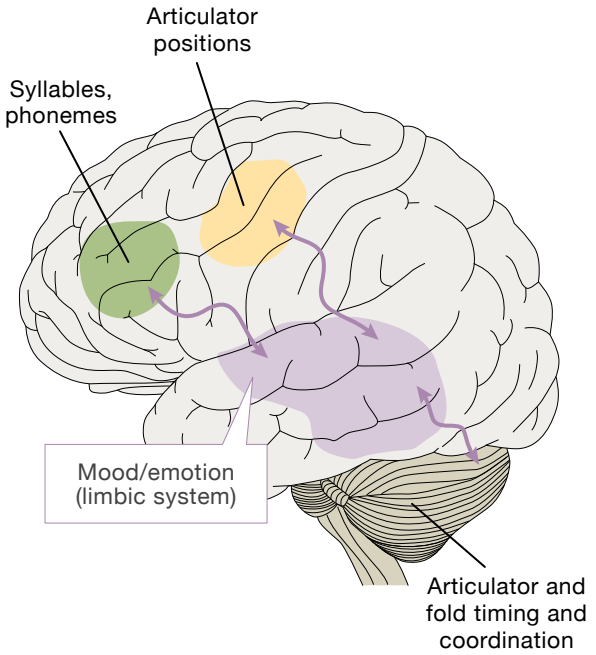


FIGURE 9. A simplified view of the modulation of representative components of the core speech network by the limbic system in depression is illustrated. Our hypothesis is that depression disrupts the limbic system, modifying core production regions required for precise timing and coordination of sound production. This hypothesis is supported by observed average slowing of speaking rate and general decline in clarity.

For our coordination-based features, Figure 10 shows examples of eigenvalue characterization from free-speech data with eigenspectra of the depressed subjects containing less power in the small eigenvalues, indicating a lower level of independent movement. This effect was observed across a spectrum of BDI scores from 83 different subject recordings in which the averages of normalized eigenvalues are plotted for different BDI score ranges (a higher score range indicates greater depression severity): 0–8 (blue), 9–19 (cyan), 20–28 (green), and 29–45 (red). Three different feature types are illustrated: vocal tract control (formants), clarity (vocal source: HNR-CPP), and facial movement (FAUs). A monotonic decrease in the average normalized eigenvalues with higher BDI score indicates that depression correlates with lower levels of independent vocal and facial movements, i.e., lowered ability for complex motor control.

For our timing features, Table 1 shows cumulative phoneme-dependent durations, obtained from estimated phoneme boundaries; for facial-based timing features, we see FAU rates obtained from their estimated posterior probabilities. Specifically, Table 1 shows the 10 highest correlations with BDI scores of average phoneme durations (a) and FAUs (b) from the AVEC dataset for speech from a read passage (North Wind) and from free speech. For facial-based timing features, we use FAU rates obtained from their estimated posterior probabilities. Notice that the aggregate measures have higher correlations than those of any individual feature type.

Our next step involved mapping the features into univariate scores that can be easily mapped into BDI predictions. To do this, we used a novel extension of a generative Gaussian mixture model (widely used for automatic speaker recognition [65]) referred to as a Gaussian staircase classifier [54], and a discriminative extreme learning machine, a single layer feedforward neural network architecture with randomly assigned hidden nodes [57].

Our overall prediction system is shown in Figure 11. In both training (construct models) and testing (apply

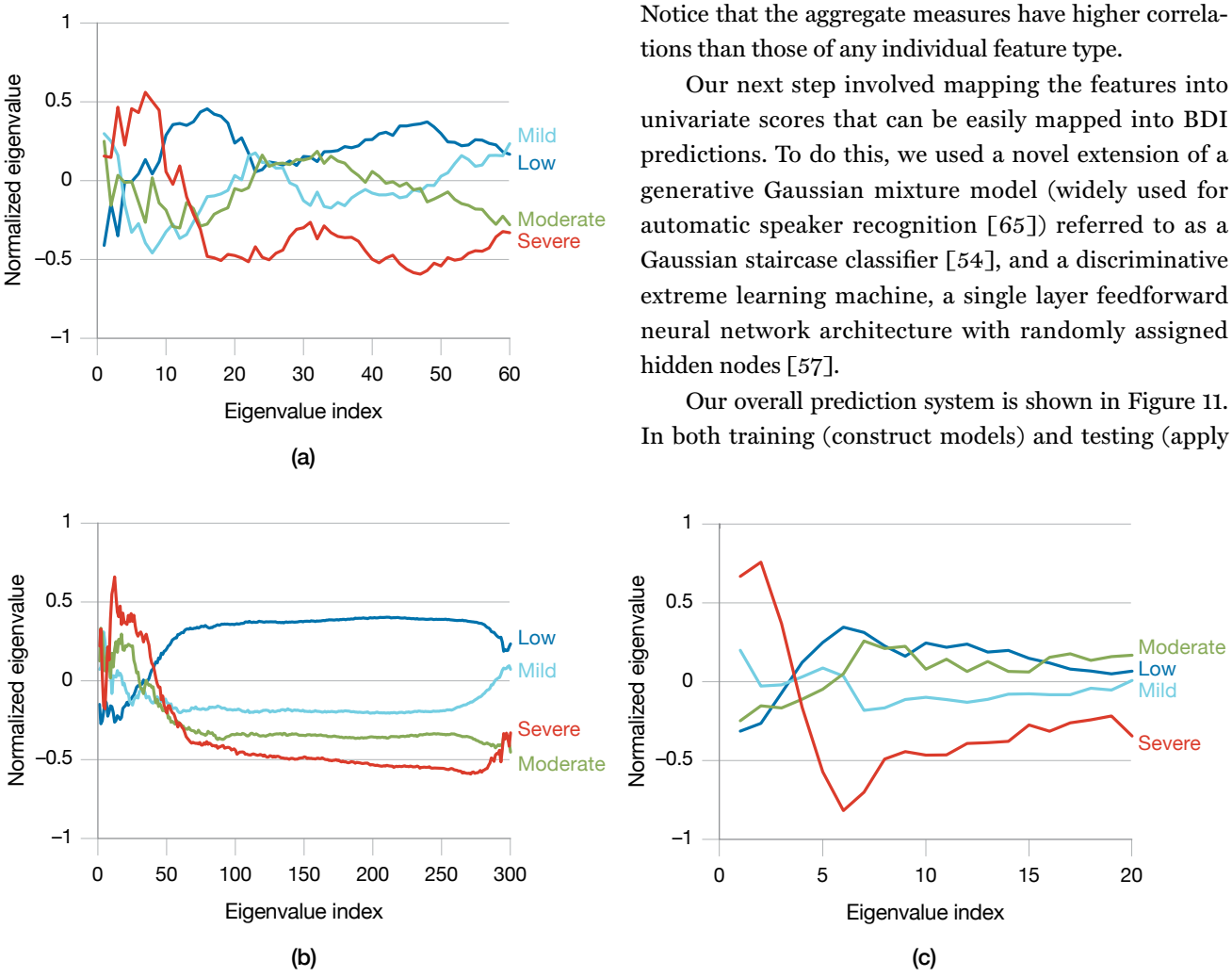


FIGURE 10. Correlation structure features are given as average normalized eigenvalues for four Beck Depression Inventory (BDI) ranges for three different feature types: (a) vocal tract control (formants), (b) facial movements (facial action units), and (c) speech clarity (vocal source: HNR-CPP). The different BDI score ranges are 0–8 (blue), 9–19 (cyan), 20–28 (green), and 29–45 (red). A higher score indicates greater depression severity.

models), after low- and high-level audio- and video-based feature extraction, we typically performed feature dimensionality reduction to account for an often high-dimensionality feature set. From the audio- and video-based features extracted, eight different feature sets were defined. Initial BDI predictions were then obtained from three predictors, which used different combinations of the eight feature sets and two types of classifiers. Each of the three predictions was obtained by using a univariate regression model created from the training set and applied to the classifier output from the test data. The outputs of the three predictors were fused to create a final BDI prediction, using weights based on each predictor's accuracy. Details of the methodology are found in Quatieri, Williamson, et al. [43] and Williamson, Quatieri, et al. [54].

The prediction system shown in Figure 11 was used for our best submission in the AVEC 2014 Challenge, with BDI score root-mean-squared error (RMSE) = 8.12 across the test data. A similar system was used in the AVEC 2013 Challenge. In both challenges, Lincoln Laboratory took first place. Note that an alternative objective is detection of depression state. We can achieve this through mapping BDI scores into two ranges—severely depressed (25–50 BDI score range) and moderately depressed (0–24)—and then mapping predicted scores to one of the two classes to form a detector. Figure 12 shows the receiver operating characteristics (ROCs) (false-alarm versus true detection rates) for this mapping in both the AVEC 2013 and 2014 Challenges.

Predicting Parkinson's Disease Severity

Parkinson's disease is a neurological disorder with associated progressive decline in motor precision and

Table 1. Correlation Coefficient (*r*) of Average Phoneme Durations (a) and Facial Action Unit (FAU) (posterior probabilities) Timing Features (b) with Beck Depression Inventory Scores

NORTH WIND		FREE SPEECH	
Phoneme	***r***	**Phoneme**	***r***
n	–0.38	ih	0.34
ah	0.35	ey	–0.27
d	–0.23	r	0.25
ih	0.22	sil	0.24
b	–0.20	oy	–0.20
g	0.18	ah	0.19
ae	0.17	y	0.16
dh	–0.15	m	0.16
eh	0.15	g	–0.14
iy	–0.15	er	0.14
Fused	0.63	Fused	0.51

(a)

	NORTH WIND		FREE SPEECH	
	FAU	***r***	**FAU**	***r***
Common FAUs for both passages	Brow lower	0.24	Brow lower	0.22
	Dimpler	–0.28	Dimpler	–0.18
	Eye widen	–0.26	Eye widen	–0.27
	Lip stretch	–0.24	Lip stretch	–0.15
	Lip tightener	–0.30	Lip tightener	–0.22
	Nose wrinkle	0.21	Nose wrinkle	0.19
	Chin raise	–0.23	Cheek raise	0.14
	Lip corner pull	–0.30	Lids tight	0.16
	Lip pucker	0.27	Lip corner depressor	0.17
	Jaw drop	0.37	Outer brow raise	–0.16
	Fused	0.58	Fused	0.46

(b)

For each session in the training set, the subject's self-reported Beck Depression Inventory assessment score was available.

sensorimotor integration, stemming presumably from the basal ganglia. In this disorder, a steady loss of cells in the midbrain leads to speech impairment in nearly 90 percent of subjects [19]. Early, accurate detection of Parkinson's disease may aid in possible intervention and rehabilitation. Thus, as with MDD, simple noninvasive biomarkers are desired for determining disease severity. Toward this end, we have applied a methodology similar to that developed for predicting MDD, introducing a novel set of acoustic speech biomarkers and fusing them with conventional features to provide clinical assessment of Parkinson's disease.

VOCAL CHARACTERIZATION

Our acoustic biomarkers reflect changes in speech production that are due to disturbances in underlying neurophysiology that affect the source, system, and prosodic components of vocal expression. In particular, impairment to the basal ganglia may in turn modulate (and cause impairment in) the core speech network, analogous to the interaction shown in Figure 9 with depression. 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 [19, 30, 31, 57, 70, 71]. Such changes occur at phonetic and larger time scales, including multiscale perturbations in formant frequency and pitch trajectories, in phoneme durations and their frequency of occurrence, and in temporal waveform structure. We have also introduced articulatory features based on a neural computational model of speech production that is introduced in more detail in the article "Fundamental Brain Research" in this issue of the *Lincoln Laboratory Journal*.

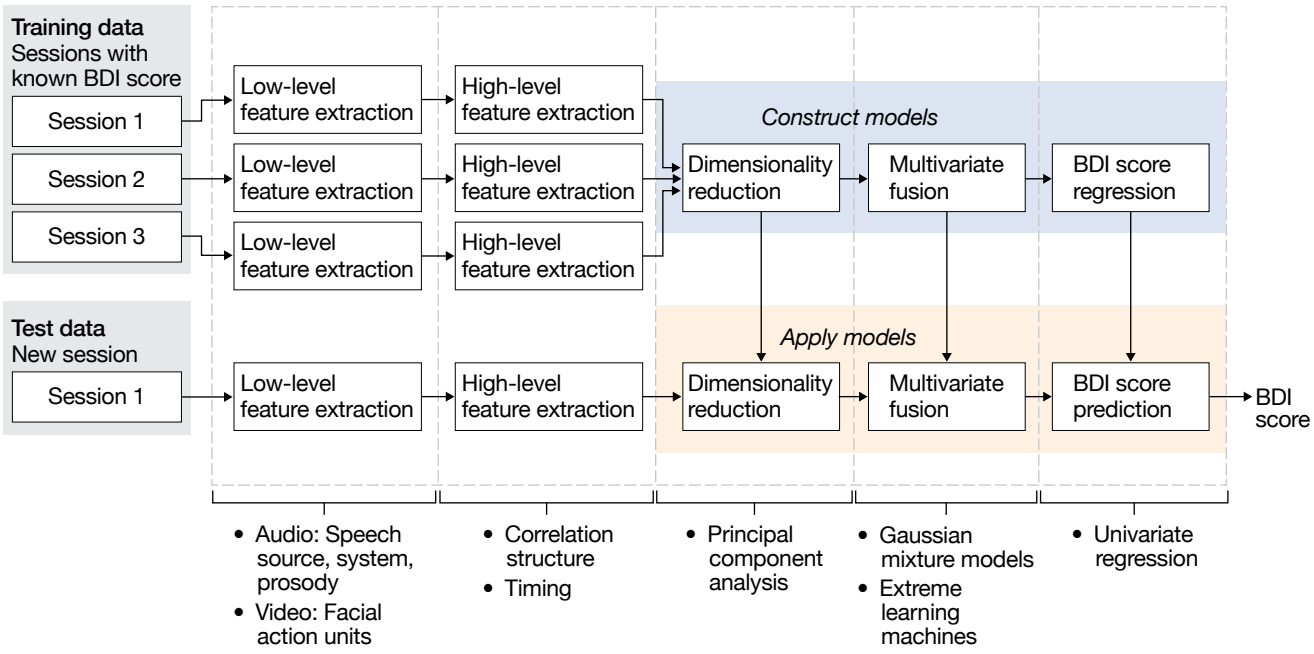


FIGURE 11. This figure shows a block diagram of the developed system for predicting the Beck Depression Inventory (BDI) assessment score. In training to obtain classifier models and in testing, the processing pipeline involves extraction of high-level features derived from low-level features, dimensionality reduction, fusion, and BDI score regression based on different classifiers. Low-level feature extraction characterizes properties of the speech or video signal within short duration (10 ms or 30 ms) time windows. High-level feature extraction computes summary statistics from these low-level features over an entire passage of recorded speech. These high-level features are composed of the eigenspectra of correlation matrices, which are created using time-delay embedding of the low-level feature time series at multiple time-delay scales. Dimensionality reduction of these eigenvalue-based feature vectors is done by using principal components analysis. Finally, Gaussian mixture models and extreme learning machines are used to produce prediction scores that are mapped into BDI predictions by using univariate regression.

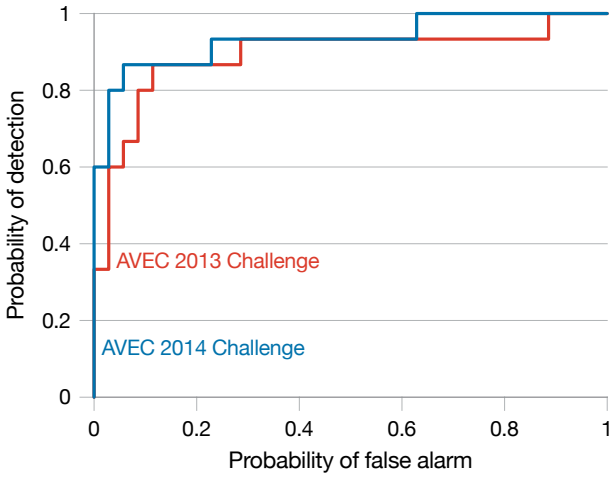


FIGURE 12. ROC detection curves (false alarm versus true detection rates) are shown for the AVEC 2013 and 2014 Challenges. In 2013, only speech was used, while in 2014, both speech and facial action units were leveraged but with considerably less training data provided. In both cases, true detection versus false-alarm rate trade-offs were favorable, i.e., for about 20 percent false alarms, we obtain about 85 percent detection accuracy. An automated system that can classify major depressive disorder at this level of accuracy would be extremely useful as a clinical screening tool.

FEATURE SELECTION

Our feature development reflects the three basic aspects of speech production: phonation (source), articulation (vocal tract), and prosody (intonation and timing). Our methodology is largely similar to those used in predicting depression severity: we design high-level features to characterize properties of timing and coordination from the low-level features. For example, our primary features exploit changes in phoneme-dependent durations, pitch slope, and formant frequency slope. The effectiveness of these features is consistent with previous findings that certain speech segments are more prone to variation than others in Parkinson's disease [72, 73].

We also use a high-level correlation structure of formant trajectories but expand this concept to a correlation structure of the position of speech articulators derived from a neurocomputational model of speech production, the Directions into Velocities of Articulators (DIVA) model [48]. (See the article "Fundamental Brain Research" in this issue for a more complete description of this approach, which uses a vocal source model in contrast to a vocal tract system model, and its application to depression.) The DIVA model takes as inputs the first three formants and the fundamental frequency of a speech utterance. Through an iterative learning process, the model then computes synaptic weights that correspond to modules of the speech production process, including aspects of the articulatory feedforward mechanism and auditory and somatosensory feedback errors (Figure 13). We hypothesize that Parkinsonian speech results from impairments in certain components of the speech production process, and therefore, when the model is trained on Parkinsonian speech via the iterative learning process, the internal variables reflect the severity of the disorder [57]. In this work, we have focused specifically on the correlation structure features derived from the DIVA model's 13 time-varying articulatory position states, with the same delay and scale parameters used for the correlation structure of the formant trajectories. More details and the complete feature selection process are described in Williamson et al. [57].

EXAMPLE DETECTION

One Parkinson's disease database we are working with is from the Interspeech 2015 Computational Paralinguistic Challenge described in Orozco-Arroyave et al. [19]. Assessments of Parkinson's severity are based on the Unified Parkinson's Disease Rating Scale (UPDRS) with a score range of 6–92 [35]. A higher score indicates higher severity. The language is Spanish and the dataset is divided into 42 tasks per speaker, yielding 1,470 recordings in the training set (35 speakers) and 630 recordings in the development set (15 speakers), both with UPDRS scores provided. The dataset 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.

As an example of the discriminatory capability of our articulatory coordination–based features derived from the DIVA neurocomputational model, Figure 14a shows eigenvalue characterization of the correlation structure features derived from the model's 13 time-varying articulatory position states. The average eigenvalues for three different ranges of Parkinson's severity are computed (in standard units) for the sentence "Luisa rey compra el colchón duro que tanto le gusta" across speakers from the training set. These averages reveal distinct differences related to Parkinson's disease severity. These

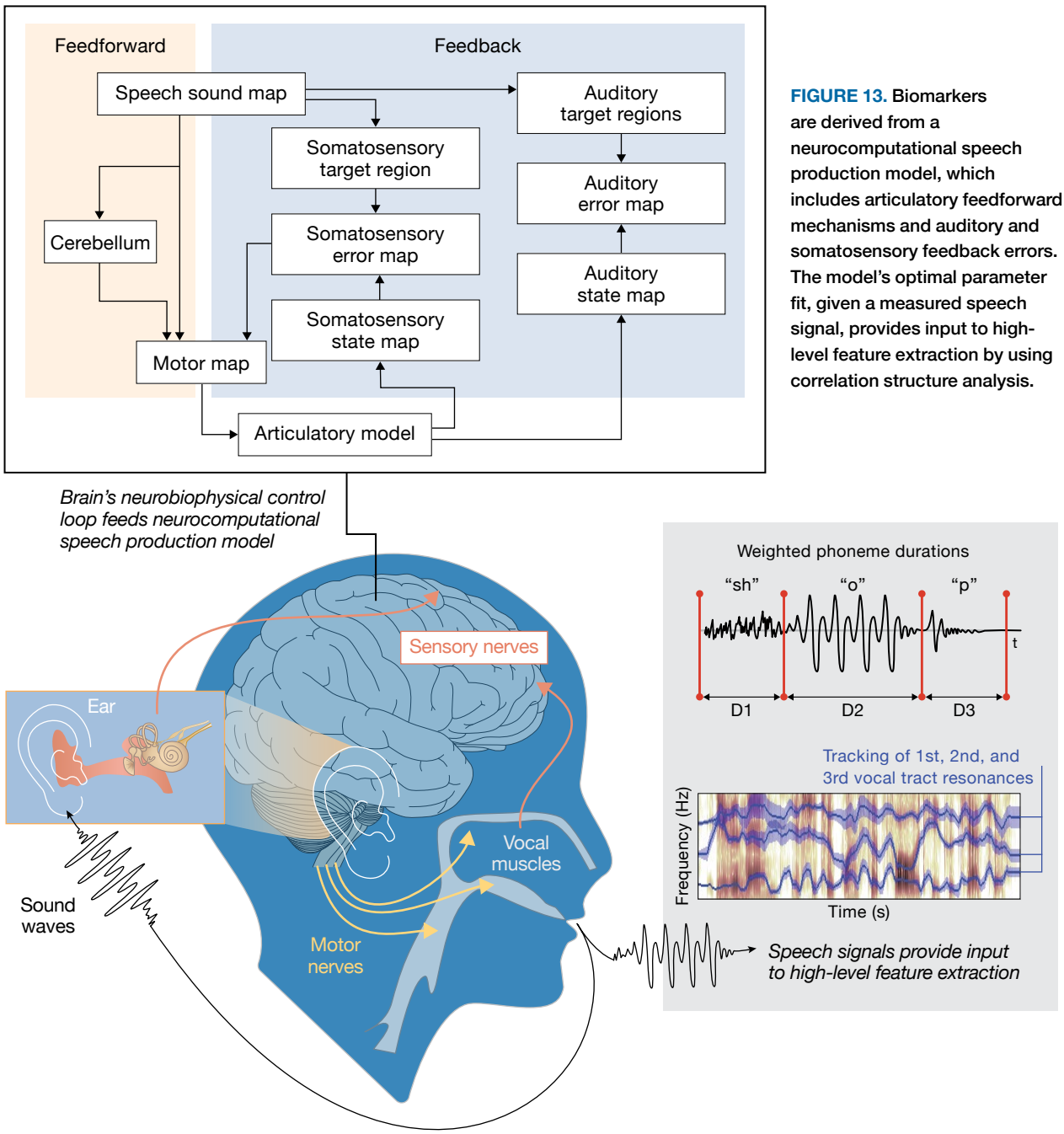


FIGURE 13. Biomarkers are derived from a neurocomputational speech production model, which includes articulatory feedforward mechanisms and auditory and somatosensory feedback errors. The model's optimal parameter fit, given a measured speech signal, provides input to high-level feature extraction by using correlation structure analysis.

differences were observed across a spectrum of scores for correlation structure features from different tasks. The decrease in the higher-indexed average normalized eigenvalues with higher Parkinson's disease score indicates that Parkinson's disease correlates with lower levels of independent vocal movements, i.e., lowered ability for complex motor control, as we observed in major depressive disorder.

In an approach similar to our detection prediction methodology, we fused a subset of our set of speech features into the Parkinson's severity prediction system by using multiple Gaussian staircase statistical regression models. Using a linear combination of the predictions from the multiple statistical models, we obtained Spearman correlations (a nonparametric measure of dependence between two variables) between predicted

scores and clinical assessments (the metric of choice for the Interspeech 2015 database) of $r = 0.63$ on the training set (fourfold cross validation), $r = 0.70$ on a held-out development set, and $r = 0.96$ on a small held-out test set [57]. As illustrated in Table 2, our neurological model–based features provide a sizable gain over our more standard timing and coordination features.

Using a system analogous to our depression prediction system, we can split the UPDRS scores in the middle and design a binary detection problem. Figure 14b shows the resulting true detection versus false-alarm rate for this binary detection scenario, indicating that the algorithm can effectively discriminate between minimal to mild and moderate to severe Parkinson's disease levels. Features used in detection are derived from a combination of primarily phonetic timing and formant coordination representations and articulatory coordination derived from the DIVA neurocomputational model [57].

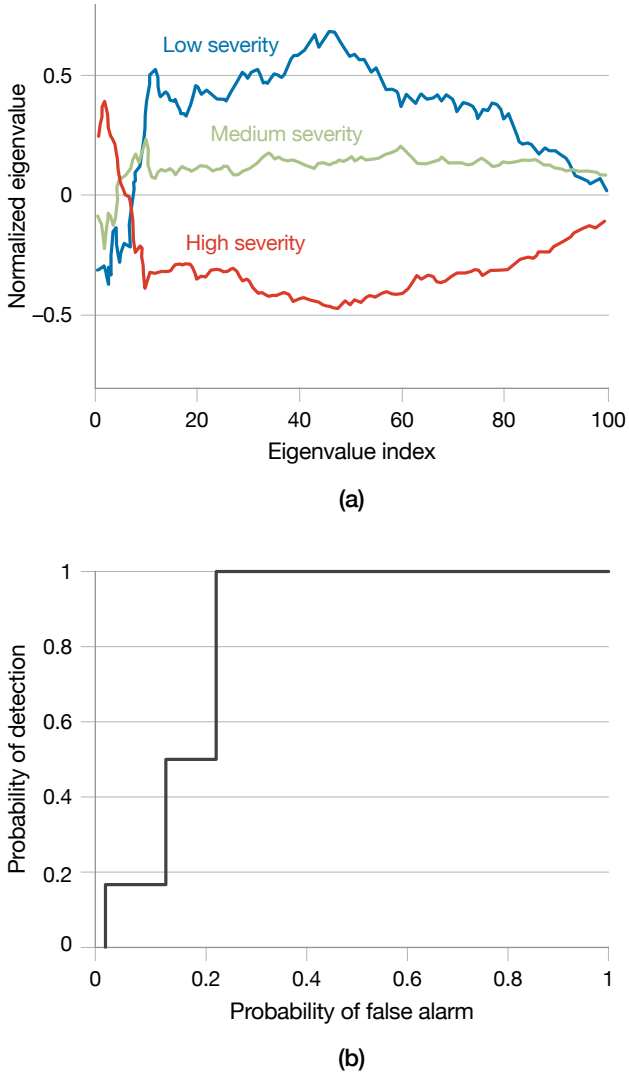


FIGURE 14. The plot in (a) shows average eigenvalue features on the development set for three ranges in the Unified Parkinson's Disease Rating Scale: 0–20 (blue), 21–40 (green), and 41–100 (red). The plot in (b) shows the true detection versus false-alarm rate for this detection scenario on the test set, indicating that the algorithm can effectively discriminate between minimal to mild and moderate to severe Parkinson's disease levels. Features used in detection are derived from a combination of primarily phonetic timing and formant coordination representations and articulatory coordination derived from the DIVA neurocomputational model [57].

Detecting Cognitive Decline from Neurotrauma

Mild traumatic brain injury (mTBI) affects an estimated 1.7 million civilians each year, who incur the condition primarily because of sport injuries, falls, and car accidents. Servicepeople are at significant risk of mTBI because of the hazards associated with military training and operations. A wide variety of sensorimotor (e.g., speech, dexterity, vestibule) and cognitive (e.g., attention, memory) problems may result from mTBI. Changes in neurocognitive status following concussion injury may adversely impact an individual's work and daily life, and, in the military context, may affect individual and unit readiness. The ability to determine cognitive performance changes associated with mTBI in an objective, noninvasive way would facilitate the monitoring of injury during all stages of care and recovery, and would serve as a valuable decision aid to leaders for determining when a soldier can safely return to duty.

Since speech and cognitive processes are highly coupled in the brain (sharing similar pathways and regions, perhaps each modulating the other), we would expect to be able to detect cognitive changes through changes in vocal and facial expression during speaking. We applied the methodology introduced earlier to a civilian case (high school athletes) and a military case (U.S. veterans), showing that our features based on timing and coordination of the complex motor activity of vocal and facial expression while speaking provide sensitive indication of cognitive impairments resulting from neurotraumatic injury. The datasets we explored represent a variety of mTBI types, some involving concussion and some requiring immediate evaluation and others requiring longer-term evaluation. Even when

Table 2. Sensitivity Improvement Plus Ability to Specify Affected Neural Pathways by Using Lincoln Laboratory Neurological Model–Based Features

	STANDARD LINCOLN LABORATORY PHYSIOLOGICAL VOCAL TRACT FEATURES	NEUROLOGICAL MOTOR COORDINATION FEATURES	STANDARD PLUS NEUROLOGICAL MOTOR FEATURES
Correlation metric	0.88	0.39	0.96

concussion does not occur, repeated exposure to head impacts without concussion can cause neurocognitive and neurophysiological impairments prior to concussion (referred to as preclinical mTBI) [37].

VOCAL AND FACIAL CHARACTERIZATION AND FEATURE SELECTION

With brain trauma, a wide range of lesion types, severities, and locations relate to different types and severities of speech and cognitive impairments. Changes in brain structure or connectivity may result in changes in source, prosodic, or articulatory aspects of voice. These changes include excessive delays in initializing a vocalization; excessive (exaggerated) or reduced (flat) emotional content; impaired function of muscles affecting the lips, tongue, vocal folds, and/or diaphragm; and difficulty in making and coordinating the precise movements of these articulators [74]. In our work, we have examined the articulatory components of speech reflected in formant tracks and the ways that changes in track dynamics and coordination map to cognitive decline. We also have focused on the change in average phoneme duration as one component of prosodic change caused by brain injury from trauma. As with formant changes, change in average phoneme duration may be associated with decline in cognitive processes from the trauma. Figure 15 shows schematically a simplified view of the link of these particular brain modules with cognitive processing that is widely distributed throughout the brain.

Our methodology is again largely similar to that used in predicting depression severity: we design high-level features to characterize properties of timing and coordination from the low-level features. Our primary features exploit changes in phoneme-dependent durations and the high-level correlation structure of formant trajectories. To characterize formant dynamics as a second

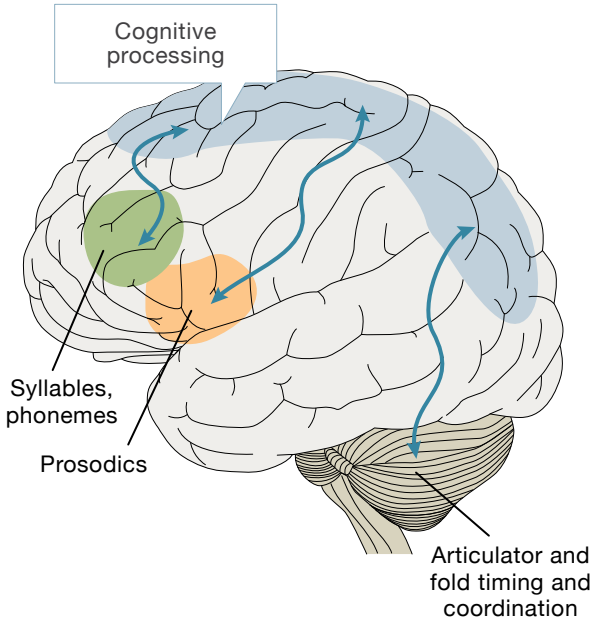


FIGURE 15. A simplified view of the representative brain modules of the core speech production network and their links to higher-level cognitive processing is illustrated. This is a highly schematized illustration given that cognitive processing is broadly distributed throughout the brain.

baseline feature, we extract nine formant functions over 20-millisecond segments at 10-millisecond 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.

DATASETS FOR CASE STUDIES

We hypothesize that preclinical or concussion-related damage results in changes in average vocal tract dynamics measured by formant frequencies,

their velocities, and acceleration; changes in articulatory coordination measured by our formant-frequency cross-correlation characterization; and phoneme-dependent average durations. These features allow machine learning algorithms to detect cognitive changes identified by a battery of cognitive tests. We developed distinct datasets for the two case studies.

1. Athlete study [55, 75]. The dataset for this study involved a population of athletes regularly receiving impacts to the head and showing signs of preclinical mTBI, a state indicated by impaired cognitive performance occurring prior to concussion. Data for this study were collected monthly in collaboration with Dr. Thomas Talavage at Purdue University under a protocol approved by an institutional review board. The study included pre-season, in-season, and post-season data from 32 high school athletes, of whom 25 were male football players and seven were female soccer players. The athletes' ages ranged from 15 to 18, with all data collected independently of any clinical diagnoses of concussions. For each athlete, the data collection included scores from the online Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT) assessment version 2.1, which comprises a series of cognitive tests typically used in the sports community [76], along with speech recordings, eye tracking, auditory perception, and optic nerve sheath-diameter measurements [75], all of which are part of a multimodal collection suite developed at Lincoln Laboratory. Although we had obtained promising preliminary results with ocular measures [75], we focused on the speech modality in the athlete study [55]. Speech features were extracted from audio recordings of the Grandfather Passage, which provides a standardized and phonetically balanced sample of speech.

   The ImPACT was used as a means of assessing cognitive performance [76]. This test is made up of six subtests that measure verbal memory, visual memory, visual motor speed, reaction time, impulse control, and a total symptom composite. For 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. A support vector machine–based classifier, with cross validation, then uses our formant features to predict a decline in cognitive performance. Figure 16a shows average z-normalized eigenspectra from formant channel-delay matrices associated with cognitive decline (red) and normal function (blue). We compared the effectiveness of vocal tract dynamics features versus articulatory coordination features. This evaluation was done via ROC curves along with a variation of their area under the curve (AUC), where a score of 1 represents a perfect test. For the reaction-time component of ImPACT, the articulatory dynamics features achieved AUC values between 0.72 and 0.98, whereas the articulatory coordination features achieved AUC values between 0.94 and 0.97. Figure 16b illustrates this comparison, indicating the importance of coordination of articulatory components over their absolute counterparts. Nevertheless, for some components of the ImPACT test, e.g., visual motor and verbal memory, the two formant-based features performed about the same.

   In this study, we also investigated features reflecting the change from baseline phoneme duration. The features were combined on the basis of their correlation with each of the cognitive modalities and then incorporated into Gaussian classifiers to predict cognitive decline. Classification performance was then analyzed using ROC curves through detection versus false alarm. Using vocal phonetic timing features for the four components of ImPACT that were studied, we computed ROC curves that demonstrated high-fidelity prediction of cognitive change. The highest AUCs achieved were 0.89, 0.80, 0.94, and 0.90 for verbal memory, visual memory, visual motor speed, and reaction time scores, respectively.
2. U.S. veterans study [77]. In this second study conducted in collaboration with Dr. Kristin Heaton at the U.S. Army Institute of Environmental Medicine (USARIEM) and Dr. Alex Lin at Brigham and Women's Hospital, data were collected from 20 subjects. Of the 20 subjects, five had a documented history of mild traumatic brain injury and 15 were control subjects. Subjects were enrolled in a larger study that used magnetic resonance spectroscopy to characterize neurochemical biomarkers for mTBI. Study participants completed a battery of cognitive

tasks selected for sensitivity to changes in cognitive performance associated with mTBI. High-quality audio and video recordings were obtained while participants completed a standardized protocol consisting of read passages, spontaneous speech, and repetition of sounds. Timing and coordination features were extracted from the speech and face data, and used along with select cognitive performance outcomes to construct statistical models for estimating cognitive performance. Speech features were based on articulatory coordination derived from the acoustic signal (vocal-tract resonances), while facial features were derived from the coordination of muscle groups (facial action units).

Gaussian staircase regression models were trained on the 15 control subjects to estimate the processing speed index (PSI) from the Wechsler Adult Intelligence Scale (WAIS-IV), which is sensitive to the cognitive consequences of mTBI. Models were then used to estimate PSI scores from the five mTBI cases not included in the training set. The strongest results were seen in predicting PSI from speech and facial features during the read passages. Pearson's correlation coefficient between the estimates and the recorded PSI scores revealed $r = 0.98$ ($p = 0.003$, $n = 5$) for the speech features and $r = 0.92$ ($p = 0.025$, $n = 5$) for the facial features. Table 3 shows these results in comparison with results for free speech.

These preliminary analyses demonstrate the promise of cognitive assessment technologies based on motor timing and coordination underlying vocal and facial expression during speaking in the context of mTBI.

In the veterans study, statistical models trained on control subjects were transferable to mTBI subjects, while the athletes study addressed accumulative injury over time. Results also show the utility of different modalities for cognitive evaluation following mTBI, with both speech and facial features affording high estimation accuracy in predicting different metrics of cognitive decline. Further validation of these technologies is required on larger datasets and in operational settings. Once validated and fully developed, these assessment technologies may provide a capability for objective, noninvasive real-time assessment of individuals in daily life and in military training and operational settings.

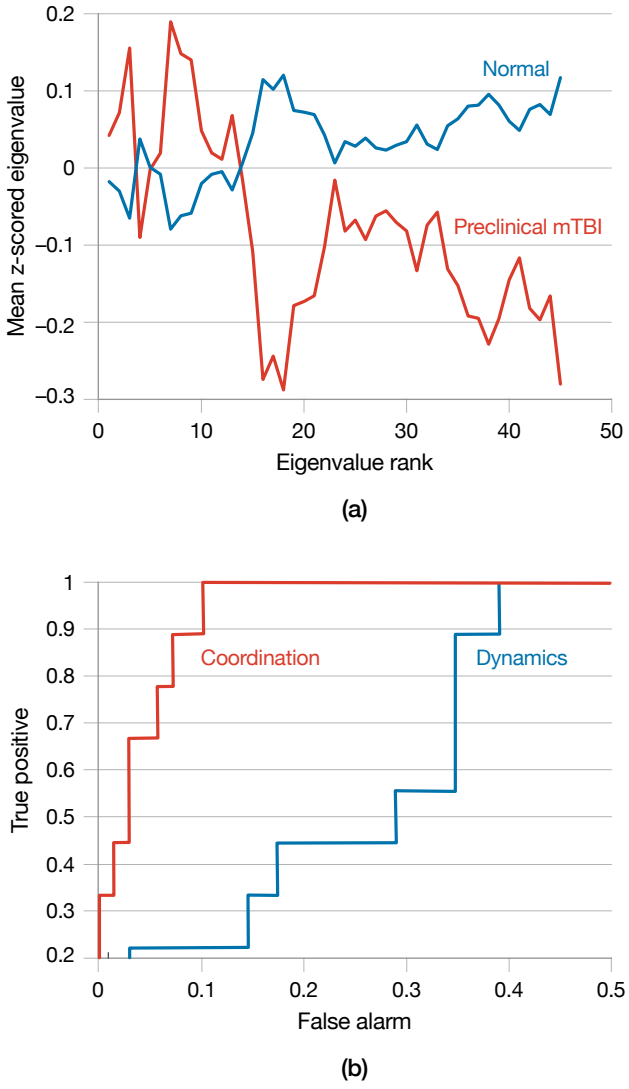


FIGURE 16. The results from the high school athlete study are shown for the reaction-time component of the ImPACT test [55]. Average z-normalized eigenspectra from formant coordination channel-delay matrices associated with cognitive decline are in red and normal function is in blue, showing a large separation especially in certain mid- and high-rank eigenvalues (a). A cognitive decline detection versus false-alarm (ROC) comparison shown in (b) gives the effectiveness of vocal tract dynamics features (blue) versus articulatory coordination features (red).

Cognitive Load

Cognitive load is often defined as the demand placed on cognitive and mental resources required by a particular task [6, 7]. Some tasks place greater burden on such resources and attributes than others. In particular, highly complex tasks, or those involving long periods of sustained activity or attention, and more monotonous

tasks requiring sustained vigilance can tax available cognitive resources and lead to fatigue. An individual's ability to adapt to changing workloads and manage fatigue caused by cognitive exertion can be influenced by stress imposed both externally (e.g., environmental extremes, physical exertion, or social interactions) and internally (e.g., psychological/emotional state, nutritional status, or changes in sleep patterns).

Applications for monitoring cognitive load, including assessing cognitive load, have been developed for both clinical and nonclinical settings. In clinical applications, the objective is often to measure the specific causes of load, while in nonclinical settings, an objective is to quickly assess cognitive ability and readiness under loaded conditions, regardless of their etiology.

Because speech and cognitive processes are highly coupled in the brain (sharing common processes and pathways, and perhaps modulating one another), we might expect to be able to determine cognitive load levels through changes in vocal and facial expression during speaking. We might think of speaking as itself a task that may occur simultaneously with some other specific targeted cognitive task of different load level. Indeed, there is evidence that in such scenarios, dual-task interference occurs under finite cognitive capacity across the dual tasks, with the relevant underlying neural mechanism located in the brain's lateral prefrontal cortex [78]. The linkage may also be due to the coordination of neuromotor activity in the brain and reliance on discrete cognitive functions in the production of speech.

Cognitive Load Detection

Our research has shown that when vocal and facial modalities are combined, they perform nearly as well as a gold-standard EEG analysis in cognitive load detection, thus providing a potential nondisruptive means to track cognitive status.

VOCAL AND FACIAL CHARACTERIZATION

On designing a multimodal database protocol that reflects typical cognitive load conditions, we employed the hypothesis that speech and the corresponding facial movements that occur while speaking are complex motor activities requiring precise neural timing and coordination. We also hypothesized that manipulating cognitive load level systematically alters this complex motor activity in a measurable way [43, 47]. This neural activation is reflected in EEG measurements, which are sometimes considered a gold standard in viewing the effect of working memory demand [11, 12, 79]. To explore those hypotheses, we designed a dual-task protocol involving auditory memory and speaking.

Table 3. Results of the U.S. Veterans Study: Predicting Processing Speed Index [77]

FEATURE SETS	READ SPEECH *r* (*p*)	FREE SPEECH *r* (*p*)
Formants	0.98 (0.003)	0.45 (0.447)
Facial Action Units	0.92 (0.025)	0.55 (0.332)

Predictions are for five mTBI subjects and from a model trained on 15 control subjects.

DUAL-TASK PROTOCOL

In one scenario, we introduced cognitive load of different levels through an auditory working memory task shown in Figure 17. Subjects gave informed consent to our working memory–based protocol approved by the MIT Committee on the Use of Humans as Experimental Subjects. Audio data were collected with a DPA acoustic lapel microphone (with a Roland Octa-Capture audio interface), facial video with a Canon high-definition video camera, and EEG signals with a 64-element Neuroscan device.

The working memory task was split into a training and a testing phase. During training, the maximum number of digits that a subject could accurately recall was estimated by using an adaptive tracking algorithm [43, 47]. This number, nc, was used to determine the three difficulty levels in the test phase, which were typically set as: dn = {ceiling(nc), ceiling(nc)-1, ceiling(nc)-2}, where ceiling corresponds to a minimum high load of 4. Despite some minor protocol changes among early subjects, this common load assessment test was used for 10 of the 11 subjects analyzed. We later defined a binary detection problem of discriminating high load (maximum number) from low load (maximum number minus two). The range of the number of digits

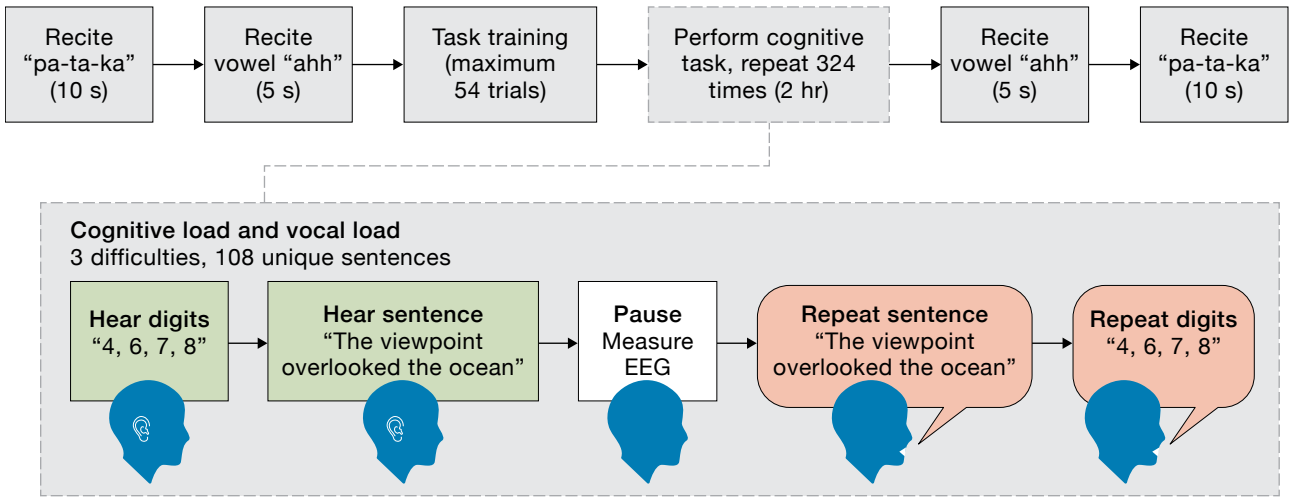


FIGURE 17. Illustrated here is a summary of the working memory protocol. The experiment consists of 324 trials. In each trial, a subject listens to a sequence of digits followed by a sentence (green boxes). Following a pause, the subject repeats the sentence and then the digits (red boxes). In different trials, the working memory load level is modified by varying the number of digits. Electroencephalogram (EEG) is measured during the pause to avoid motion and muscle artifacts. Speech (from audio and facial video) is measured when the sentence is repeated. The working memory load level is estimated based on the EEG and speech signals. Prior to and after the two-hour cognitive task, we introduced reciting the diadochokinetic sequence "pa-ta-ka" and the vowel "ahh," allowing for analysis of speech samples pre- and post-fatigue that are standardly used in articulatory and vocal source analysis.

recalled across all subjects was 2–5 for low load and 4–7 for high load.

Finally, with our protocol, we also measured skin conductance, temperature, and pulse oxygenation level. Though not our focus, the study of these measurements may pave the way to investigating other multimodal biomarkers.

FEATURE SELECTION

Our features from three modalities (voice, face, EEG) are based on the principles of timing and coordination within components of each modality. In each case, we extracted first baseline low-level features followed by our high-level features as functions of the low-level features. For voice, feature vectors were extracted only from the single spoken sentence component of each trial in the test phase of the auditory memory task. Low-level vocal features comprise measures of phoneme and pseudosyllable durations, pitch dynamics, spectral (formant) dynamics, and vocal-fold irregularity (creak). We constructed high-level features that capture timing and interrelationships across the low-level features. The feature sets were derived under the hypothesis that differences in cognitive load produce detectable changes in speech production timing and coordination within and across articulatory and vocal-fold components. For facial expression, analyzed during the same time interval as audio, the extracted low-level features were facial action units [22, 24, 27], followed by correlation-based measures as high-level features. For the EEG, during the pause interval to avoid motion and muscle artifacts, we performed preprocessing to extract low-level EEG signals free of many typical artifacts, followed by correlation and frequency-dependent coherence and power measures. To avoid motion and muscle artifacts, the EEG measurements were made during time intervals when the subject was not speaking.

As was done in the research on neurological conditions, we averaged across different targeted class conditions. Our high-level features showed strong load discriminability with averages across all subjects of normalized (z-scored) eigenvalues from formant, creak, and delta-MFCC signals for low load and high load. In all three cases, there was greater power in the medium-level eigenvalues during higher cognitive load. This finding indicates greater dynamical complexity in formant frequencies, creak, and spectral content during higher cognitive load and thus higher levels of independent

vocal movements than in the low-load case. Although this finding corresponds to more complex motor control under high load, we hypothesize that the eigenvalue distributions indicate greater independence associated with more random or erratic movements under the high load state—in contrast to our earlier observations for MDD, Parkinson's disease, and mTBI, in which we found less independence of vocal movements. In our cognitive load experiments, we found a similar level of discriminability based on class-conditioned averages of the EEG high-level features, which are coherence eigenvalues of spatial log-power values in the beta frequency band. Also, high load is associated with lower levels of EEG power.

EXAMPLE DETECTION

Although our protocol involves feature processing of single spoken sentences, the ability to detect load after fusing evidence across multiple sentences can be assessed by combining the Gaussian classifier scores from different trials, provided that the trials involve the same load condition. This combining was done by randomly selecting, from the same subject, a number of trials of either high load or low load and summing their Gaussian classifier scores. For each subject, load condition, and combination number, 200 randomly chosen sets of trials were used to determine the fused scores across multiple sentences.

Figure 18 summarizes the ROC results (detection versus false alarm) for the modalities in various combinations. We observe that the EEG-based detector rapidly converges to a limit that, after six minutes, slightly outperforms the combined audio and video modalities that converge to a limit more slowly. Each modality alone (audio or video) converges to a limit that underperforms the combination of audio and video, with audio outperforming video [43] (not shown). Observe that combining all three modalities provides only a small gain over the EEG.

Predicting Cognitive Fatigue

Individuals often report the subjective impression of weakness or slowness in performing cognitive tasks, particularly following periods of heavy cognitive workload. This subjective feeling has been called cognitive fatigue and is closely associated with the objective notion of cognitive capacity (alluded to earlier in the context of dual tasks), which is the ability of an individual to perform cognitive work. When cognitive workload exceeds an individual's overall capacity, at a given point in time, performance in managing the workload can be expected to degrade. For individuals engaged in tasks such as driving a car or piloting an airplane, reductions in cognitive capacity may lead to accidents and injuries. The ability to monitor changes in cognitive capacity over time would have benefits for planning and decision making in both civilian and military settings, and may inform interventions and performance-enhancement techniques.

A frequently used method for assessing cognitive capacity in real time is the psychomotor vigilance task (PVT), which is a test that measures reaction time and attentiveness, developed primarily as an in-laboratory assessment. Despite recent efforts to integrate PVT into portable and/or wearable devices, the test itself is still largely inappropriate for operational environments because it requires the tested individuals to disengage from their primary task and spend between two and 20 minutes participating in the PVT. Some success has been achieved in developing nonobtrusive assessment methods that are based on ocular behavior, such as the percentage of eyelid closure (e.g., PERCLOS) and ocular dynamics (e.g., SmoothEye). These methods hold promise for monitoring cognitive capacity in field settings but nonetheless require specialized equipment and/or specific environmental conditions to function accurately.

VOICE CHARACTERIZATION AND FEATURE SELECTION

In ongoing work with the U.S. Army and U.S. Air Force, we have been investigating voice-based technologies for real-time assessment of cognitive capacity. We are using a variety of standard acoustic speech features in combination with the timing and coordination features developed by Lincoln Laboratory. Voice-based assessments for cognitive capacity are attractive because they can be made nonobtrusive, can be implemented in hardware that uses conventional microphones, and have the potential to be adapted to standard-issue communications platforms used in operational environments, such as airplane cockpits. Early development and validation have been taking place on data collected at Lincoln Laboratory. Data collections with U.S. Army and U.S. Air Force partners are ongoing and will substantially advance early results and validation efforts.

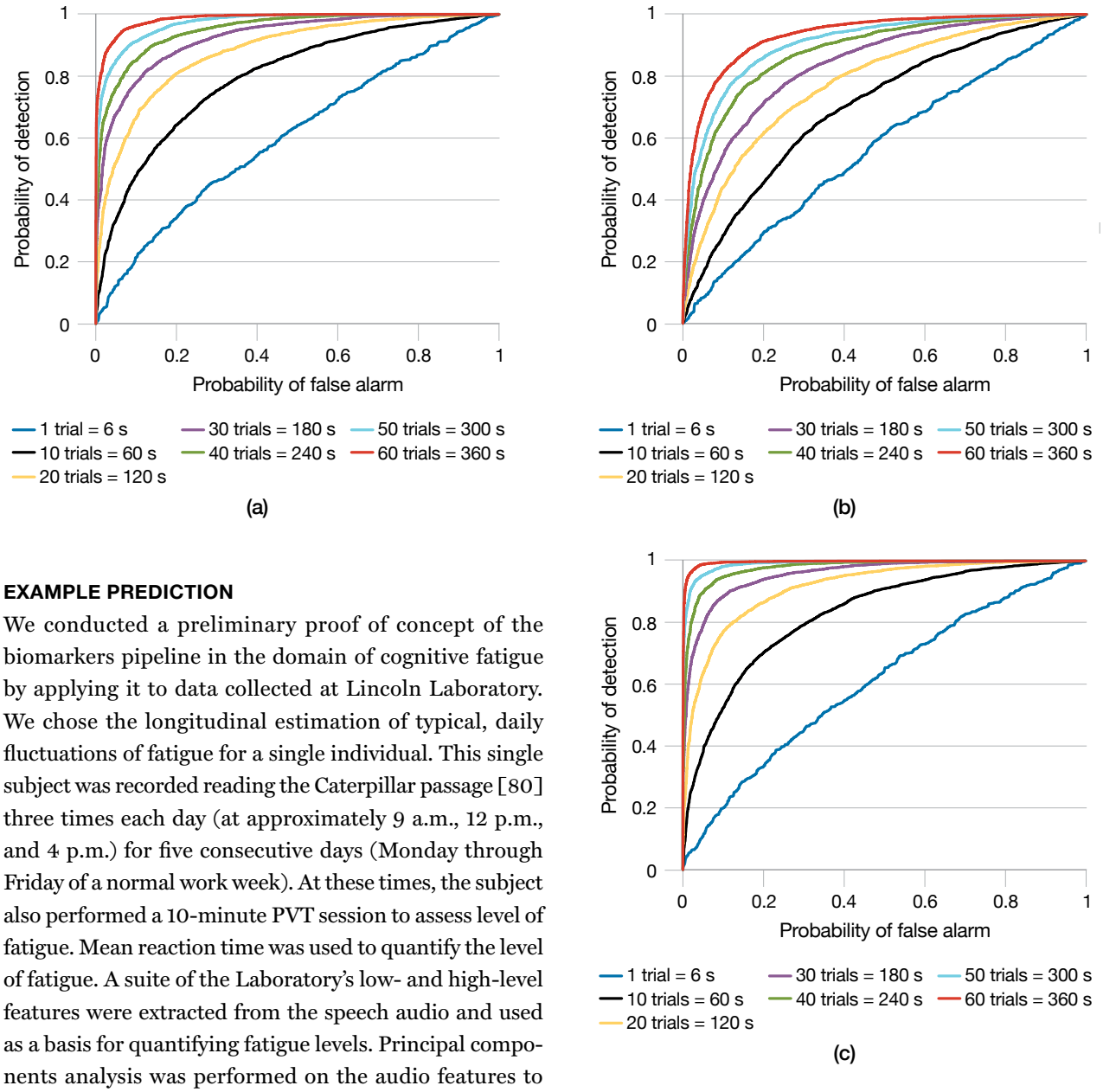


FIGURE 18. The fused probability of detection versus false alarm for the modalities of EEG (a), audio and video (b), and EEG, audio, and video (c) is compared. Each panel gives ROCs as a function of increasing number of trials from 1 to 60, corresponding to 6 seconds to 360 seconds (6 minutes) for low and high cognitive loads.

EXAMPLE PREDICTION

We conducted a preliminary proof of concept of the biomarkers pipeline in the domain of cognitive fatigue by applying it to data collected at Lincoln Laboratory. We chose the longitudinal estimation of typical, daily fluctuations of fatigue for a single individual. This single subject was recorded reading the Caterpillar passage [80] three times each day (at approximately 9 a.m., 12 p.m., and 4 p.m.) for five consecutive days (Monday through Friday of a normal work week). At these times, the subject also performed a 10-minute PVT session to assess level of fatigue. Mean reaction time was used to quantify the level of fatigue. A suite of the Laboratory's low- and high-level features were extracted from the speech audio and used as a basis for quantifying fatigue levels. Principal components analysis was performed on the audio features to achieve dimensionality reduction, keeping the first five most prominent components. A linear model was trained with a leave-one-session-out cross validation to predict mean reaction time from the audio features. Results in Figure 19 show a strong correlation ($r^2 = 0.64$, $p < 0.01$) between actual and estimated reaction time, with a root-mean-square error (RMSE) of 9.6 milliseconds, or approximately 15.5 percent of the observed range of mean reaction times in the dataset. These results, though preliminary, demonstrate the promise of vocal biomarkers for cognitive fatigue assessment and, more generally, for cognitive status monitoring. The ability to predict objective measures of fatigue with an estimation error of approximately 16 percent may be relevant and actionable for individuals on sleep-restricted schedules during which fatigue levels have been shown to vary by 35 percent or more [81]. Preliminary results on larger datasets have revealed improved performance compared

to this modest-sized proof of concept. Our expanded effort in this area is also investigating the independent and joint effects of cognitive load and fatigue conditions in voice features, showing the changing discriminatory ability of cognitive load under different cognitive fatigue conditions [82].

Other Modalities

Lincoln Laboratory is also expanding to other behavioral measures, including fine motor (e.g., eye tracking and hand dexterity) and gross motor movements (e.g., static and dynamic balance), physiological measures (e.g., heart rate and skin conductance and temperature), and brain-computer interfaces (e.g., EEG measures), along with corresponding neurobiophysical models, for detecting, phenotyping, and monitoring brain trauma and disease. For example, eye tracking is included in our high school athlete data collection for preclinical mTBI in our collaboration with Purdue University and also in a cognitive load data collection in collaboration with the Combat Capabilities Development Command Soldier Center (CCDC Soldier Center), formerly the U.S. Army Natick Soldier Research, Development, and Engineering Center. In each case, eye-tracking features have contributed significantly to the prediction of cognitive decline [75] and cognitive load level [83], respectively. Physiological measures also play a role in our past and ongoing investigations of depression (as part of our multimodal study with the Wyss Institute at Harvard Medical School) and of cognitive overload and fatigue (as part of our multimodal studies with USARIEM and CCDC Soldier Center). In addition to using EEG analysis in cognitive load level prediction, we are using it to obtain insights into the importance of active brain regions during visual and auditory memory tasks [83]. Finally, in this same context of cognitive load prediction, we are investigating the effect of movement in cognition. One example involves recording speech, EEG (64-channel Neuroscan system), and full-body motion capture (17-accelerometer Xsens System) from 10 subjects in both standing and walking conditions while they are engaged in an auditory working memory task. Here, we are exploring the effect of cognitive load on timing and coordination features within and across the three modalities, their interrelationships, and the consequence of multitasking [84].

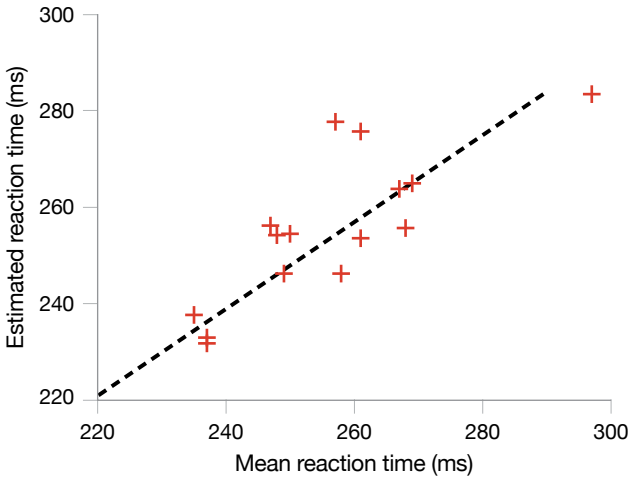
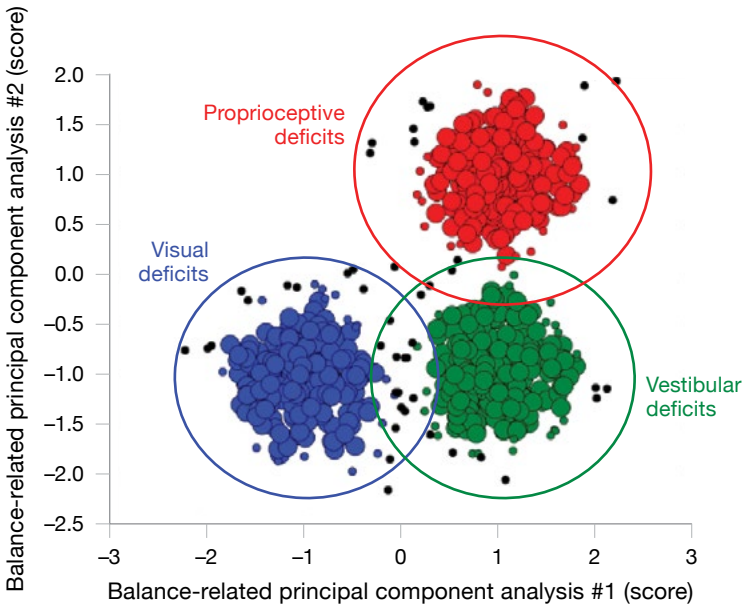


FIGURE 19. Preliminary results of using speech-based biomarkers for predicting reaction time are plotted. Reaction time, as measured with the psychomotor vigilance task, is sometimes taken as a gold standard for assessment of alertness and cognitive fatigue. A linear model was trained with a leave-one-session-out cross validation to predict reaction time from speech features, using data collected at Lincoln Laboratory of a single subject assessed three times daily for one work week.

Providing sensitive detection simultaneously with specific phenotyping is an inherent challenge facing the assessment of many different neurological conditions. An especially prevalent example of these challenges can be seen in the domain of mTBI, where a wide variety of sensorimotor (e.g., vestibular deficits) and cognitive (e.g., attention deficits) impairments lead to subtle and heterogeneous signs and symptoms. Clinical assessments in this domain tend to be subjective and insensitive, and neural imaging is expected to be negative. Moreover, diagnosis is typically focused on classifying the injury into broad categories: mild, moderate, or severe. There is a pressing need for technologies that can provide individualized physiological phenotypes in order to facilitate diagnostic specificity, which is often complicated by the co-occurrence of conditions with similar symptom profiles (e.g., PTSD and depression), as well as to tailor treatment approaches to individual needs and more accurately monitor risk and post-injury recovery.

In one program, Lincoln Laboratory staff—in collaboration with clinical and domain experts from Spaulding Rehabilitation Hospital, Massachusetts General Hospital, and Harvard Medical School—are developing

(a)



(b)

FIGURE 20. Neurocomputational model–based approaches to phenotyping mTBI are being developed in the virtual reality environment at the Laboratory's Sensorimotor Technology Realization in Immersive Virtual Environments (STRIVE) Center (a). Behavioral and physiological responses to sensorimotor provocative tests are measured, driving neurocomputational models of motor control in normal and impaired individuals. This testing can facilitate system identification of impairment in neurological control mechanisms involving vestibular, visual, and proprioceptive sensory feedback (b).

methods for individualized phenotyping of mTBI-related impairments through the analysis of balance and gait. Balance and gait show frequent impairment in mTBI from a variety of causes (e.g., vestibular, visual-motor, and/or proprioceptive sensory feedback impairment) and therefore hold promise for enabling phenotyping. Two parallel approaches are underway (see Figure 20). The first is the development of sensorimotor provocative tests, conducted on a flexible treadmill platform in the virtual reality environment at the Laboratory's Sensorimotor Technology Realization in Immersive Virtual Environments (STRIVE) Center. These tests are designed to bring out latent impairments not observable in activities of daily living. The second approach

is the development of neurocomputational models of motor control in normal and impaired individuals to facilitate system identification of neurological control mechanisms. These approaches are synergistic, with data from STRIVE Center experiments used to refine control models and models used to inform the design of sensorimotor provocative tests.

Long-Term Vision

In the Laboratory's Human Health and Performance Systems Group, we have previous or ongoing projects in detecting and tracking amyotrophic lateral sclerosis (ALS) [84], dementia [86], and autism spectrum disorder [87]. A common theme to detecting and tracking these conditions is the discovery of behavioral-based biomarkers that reflect a change or decline in brain functioning as manifested in motor control, and more specifically changes in timing and coordination within and across components of behaviors. It is hypothesized that these relations are associated with neural coordination across different parts of the brain that are essential in motor control.

As a final application, we mention our introduction into an area of immediate urgency: detecting and tracking COVID-19 (the novel coronavirus) through asymptomatic and symptomatic stages. Given the physiologically based insult to breathing functions [88] and the growing evidence of neurological deficits present in COVID-19 [89, 90], we hypothesize that biomarkers derived from measures of vocal-subsystem coordination that includes both lower and upper respiratory systems may provide a sensitive indicator of COVID-19, most importantly in its asymptomatic stages. Preliminary results with audio interviews of two subjects reveal strong effect sizes in distinguishing pre-COVID-19 (pre-exposure) from post-COVID-19 (after positive diagnosis but still asymptomatic) by using intensity of breathing (respiration during speech), coordination of respiration and pitch (fundamental frequency), and coordination of pitch and articulatory motion. Morphology of eigenvalue Cohen's *d* effect sizes indicates a constricted breathing and reduced complexity of coordinated subsystem movement. Although preliminary results are promising, we clearly need to validate these results and address confounding influences with larger, more controlled datasets.

With our timing and coordination-based features, we have achieved effective detection across a variety of application areas. Nevertheless, to more strongly validate performance in all areas, we need to move from the lab and clinic to the field. To provide access to larger populations, we are currently translating our data collections and detection algorithms to apps using mobile technology, e.g., smartphones and tablets. In the area of depression, we are working with the commercial entity Sonde Health; in traumatic brain injury, we are collaborating with the U.S. Army Medical Materiel Development Activity (USAMMDA) to develop a mobile device approved by the U.S. Food and Drug Administration; and in cognitive load and fatigue, we are collaborating with USARIEM toward a predictive algorithm for assessment within a technology transfer agreement. Toward our goal of early warning and tracking of COVID-19, it will be essential to address potential confounders, such as different recording environments and channels, unbalanced data quantities, and changes in underlying vocal status from pre-COVID-19 exposure to post-COVID-19 diagnosis. Finally, it will be essential to understand the specificity of our proposed biomarkers; for example, these proposed biomarkers must be able to differentiate COVID-19 from the typical flu and flu-like conditions resulting in various forms of inflammation.

A growing path in our biomarker research and development involves pairing our behavioral approaches with neurocomputational biophysical modeling and with clinical observation. This trifold strategic approach fuses empirical measurements, phenomenological (data-driven) models, and mechanistic models.

Our long-term vision is to move away from predictions based on group analysis and to use the trifold fusion to improve assessments of mechanisms underlying individual impairments, leading to personalized design of patient care monitoring and intervention systems. Our ultimate objective is to develop patient-specific neurocomputational models of human behavior that integrate sensorimotor impairments for early warning and improved specificity and phenotyping.

Acknowledgments

The authors acknowledge for discussions, assistance, and support Laura Brattain, Gregory Ciccarelli, Nicolas Malyska, Daryush Mehta, Tejash Patel, and Christopher

Smalt from Lincoln Laboratory; Satrajit Ghosh from the McGovern Institute for Brain Research at MIT; Thomas Talavage from Purdue University; Alexander Lin from Brigham and Women's Hospital; and Marianna Eddy and Joseph Moran from the U.S. Army. ■

References

1. World Health Organization, Mental Health, https://www.who.int/mental_health/en/.
2. Office of Suicide Prevention, "Suicide Among Veterans and Other Americans," U.S. Department of Veterans Affairs, 3 Aug. 2016, available at http://www.mentalhealth.va.gov/docs/2016suicidedatareport.pdf.
3. World Health Organization, "Dementia Fact Sheet," available at http://www.who.int/en/news-room/fact-sheets/detail/dementia.
4. Centers for Disease Control and Prevention, "Community Report on Autism 2018," U.S. Department of Health and Human Services.
5. T. Elianan and L.H. Joycox, eds., *Invisible Wounds of War: Psychological and Cognitive Injuries, Their Consequences, and Services to Assist Recovery*. Santa Monica: RAND Corporation, 2008.
6. S.E. Lively, D.B. Pisoni, W. Van Summers, and R.H. Bernacki, "Effects of Cognitive Workload on Speech Production: Acoustic Analyses and Perceptual Consequences," *Journal of the Acoustical Society of America*, vol. 93, no. 5, 1993, pp. 2962–2973.
7. B. Yin, F. Chen, N. Ruiz, and E. Ambikairajah, "Speech-Based Cognitive Load Monitoring System," paper in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2008, pp. 2041–2044.
8. B. Yin and F. Chen, "Towards Automatic Cognitive Load Measurement from Speech Analysis," paper in *Proceedings of the 12th International Human-Computer Interaction Conference, Interaction Design and Usability*, 2007, pp. 1011–1020.
9. M.A. Khawaja, N. Ruiz, and F. Cheng, "Think Before You Talk: An Empirical Study of Relationship between Speech Pauses and Cognitive Load," paper in *Proceedings of the 20th Australasian Computer-Human Interaction Conference*, 2008, pp. 335–338.
10. P. Le, J. Epps, H.C. Choi, and E. Ambikairajah, "A Study of Voice Source- and Vocal Tract-Based Features in Cognitive Load Classification," paper in *Proceedings of the International Conference on Pattern Recognition*, 2010, pp. 4516–4519.
11. P. Zarjam, J. Epps, and F. Chen, "Characterizing Working Memory Load Using EEG Delta Activity," paper in *Proceedings of the 19th European Signal Processing Conference*, 2011, pp. 1554–1558.
12. P. Zarjam, J. Epps, and F. Chen, "Evaluation of Working Memory Load Using EEG Signals," paper in *Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, 2011, pp. 715–719.
13. N. Cummins, S. Scherer, J. Krajewski, S. Schnieder, J. Epps, and T.F. Quatieri, "A Review of Depression and Suicide Risk Assessment using Speech Analysis," *Speech Communication*, vol. 17, 2015, pp. 10–49.
14. E. Moore, M. Clements, J. Peifer, and L. Weisser, "Analysis of Prosodic Variation in Speech for Clinical Depression," paper in *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2003, pp. 2925–2928.
15. J.C. Mundt, P.J. Snyder, M.S. Cannizzaro, K. Chappie, and D.S. Geralts, "Voice Acoustic Measures of Depression Severity and Treatment Response Collected via Interactive Voice Response (IVR) Technology," *Journal of Neurolinguistics*, vol. 20, no. 1, 2007, pp. 50–64.
16. A. Ozdas, R.G. Shiavi, S.E. Silverman, M.K. Silverman, and D.M. Wilkes, "Investigation of Vocal Jitter and Glottal Flow Spectrum as Possible Cues for Depression and Near-Term Suicidal Risk," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 9, 2004, pp. 1530–1540.
17. J.K. Darby, N. Simmons, and P.A. Berger, "Speech and Voice Parameters of Depression: A Pilot Study," *Journal of Communication Disorders*, vol. 17, no. 2, 1984, pp. 75–85.
18. J.F. Greden and B.J. Carroll, "Psychomotor Function in Affective Disorders: An Overview of New Monitoring Techniques," *The American Journal of Psychiatry*, vol. 138, no. 11, 1981, pp. 1441–1448.
19. J.R. Orozco-Arroyave, J.D. Arias-Londoño, J.F. Vargas-Bonilla, M.C. González-Rátiva, and E. Nöth, "New Spanish Speech Corpus Database for the Analysis of People Suffering from Parkinson's Disease," paper in *Proceedings of the Ninth International Conference on Language Resources and Evaluation*, 2014, pp 342–347.
20. P. Wang, F. Barrett, E. Martin, M. Milonova, R.E. Gur, R.C. Gur, et al., "Automated Video-Based Facial Expression Analysis of Neuropsychiatric Disorders," *Journal of Neuroscience Methods*, vol. 168, no. 1, 2008, pp. 224–238.
21. M. Valstar, B. Schuller, K. Smith, T. Almaev, F. Eyben, J. Krajewski, et al., "AVEC 2014: 3D Dimensional Affect and Depression Recognition Challenge," paper in *Proceedings of the Fourth International Workshop on Audio/Visual Emotion Challenge*, 2013, pp. 3–10.
22. W. Gaebel and W. Wölwer, "Facial Expression and Emotional Face Recognition in Schizophrenia and Depression," *European Archives of Psychiatry and Clinical Neuroscience*, vol. 242, no. 1, 1992, pp. 46–52.
23. G.E. Schwartz, P.L. Fair, P. Salt, M.R. Mandel, and G.L. Klerman, "Facial Expression and Imagery in Depression: An Electromyographic Study," *Psychosomatic Medicine*, vol. 38, no. 5, 1976, pp. 337–347.
24. P. Ekman, W.V. Freisen, and S. Ancoli, "Facial Signs of Emotional Experience," *Journal of Personality and Social Psychology*, vol. 39, no. 6, 1980, pp. 1125–1134.

25. P. Zarjam, J. Epps, and N.H. Lovell, "Beyond Subjective Self-Rating: EEG Signal Classification of Cognitive Workload," *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 4, 2015, pp. 301–310.
26. M. Fava and K.S Kendler, "Major Depressive Disorder," *Journal of Neuroscience*, vol. 28, no. 2, 2000, pp. 335–341.
27. G. Littlewort, J. Whitehill, T. Wu, I.R. Fasel, M. Frank, J.R. Movellan, and M.S. Bartlett, "The Computer Expression Recognition Toolbox (CERT)," paper presented at the Ninth IEEE International Conference on Automatic Face and Gesture Recognition, Santa Barbara, 21–23 Mar. 2011.
28. 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, 2012, pp. 1732–46.
29. A.C. Trevino, T.F. Quatieri, and N. Malyska, "Phonologically-Based Biomarkers for Major Depressive Disorder," *EURASIP Journal on Advances in Signal Processing*, vol. 2011, no. 1, 2011, pp. 1–18.
30. 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, 1963, pp. 221–229.
31. 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, 1978, pp. 47–57.
32. 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, 2012, pp. 1208–1219.
33. S. Skodda and U. Schlegel, "Speech Rate and Rhythm in Parkinson's Disease," *Movement Disorders*, vol. 23, no. 7, 2008, pp. 985–992.
34. J. Rusz, R. Cmejla, T. Tykalova, H. Ruzickova, J. Klempir, V. Majerova, et al., "Imprecise Vowel Articulation as a Potential Early Marker of Parkinson's Disease: Effect of Speaking Task," *The Journal of Acoustical Society of America*, vol. 134, no. 3, 2013, pp. 2171–2181.
35. C.G. Goetz, B.C. Tilley, S.R. Shaftman, G.T. Stebbins, S. Fahn, P. Martinez-Martin, et al., "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, 2008, pp. 2129–2170.
36. M. Falcone, N. Yadav, C. Poellabauer, and P. Flynn, "Using Isolated Vowel Sounds for Classification of Mild Traumatic Brain Injury," paper in *Proceedings of the 2013 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2013, pp. 7577–7581.
37. T.M. Talavage, E.A. Nauman, E.L. Breedlove, U. Yoruk, A.E. Dye, K.E. Morigaki, et al., "Functionally-Detected Cognitive Impairment in High School Football Players without Clinically-Diagnosed Concussion," *Journal of Neurotrauma*, vol. 31, no. 4, 2014, pp. 327–338.
38. C. Cao, R.L. Tutwiler, and S. Slobounov, "Automatic Classification of Athletes with Residual Functional Deficits Following Concussion by Means of EEG Signal Using Support Vector Machine," *IEEE Transactions of Neural Systems and Rehabilitation and Engineering: A Publication of the IEEE Engineering in Medicine and Biology Society*, vol. 16, no. 4, 2008, pp. 327–335.
39. C. Poellabauer, N. Yadav, L. Daudet, S.L. Schneider, C.B. Busso, and P. Flynn, "Challenges in Concussion Detection Using Vocal Acoustic Biomarker," *IEEE Acces*s, vol. 3, 2015, pp. 1143–1160.
40. A.L. Russo, J. Verghese, A.L. Metti, R.M. Boudreau, H.J. Aizenstein, S. Kritchevsky, et al., "Slowing Gait and Risk for Cognitive Impairment: The Hippocampus as a Shared Neural Substrate," *Neurology*, vol. 89, no. 4, 2017, pp. 336–342.
41. U.U. Samadani, R. Ritlop, M. Reyes, E. Nehrbass, M. Li, E. Lamm, et al., "Eye Tracking Detects Disconjugate Eye Movements Associated with Structural Traumatic Brain Injury and Concussion," *Journal of Neurotrauma*, vol. 32, 2015, pp. 548–556.
42. E. Gowen and A. Hamilton, "Motor Abilities in Autism: A Review Using a Computational Context," *Journal of Autism and Development Disorders*, vol. 43, no. 2, 2013, pp. 323–344.
43. T.F. Quatieri, J.R. Williamson, C.J. Smalt, J. Perricone, T. Patel, L.J. Brattain, et al., "Multi-modal Biomarkers to Discriminate Cognitive State," *The Role of Technology in Clinical Neuropsychology*. Eds. R.L Kane and T.D. Parsons. New York: Oxford University Press, 2017.
44. 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," paper in *Proceedings of the 15th Annual Conference of the International Speech Communication Association*, 2014, pp. 1038–1042.
45. B. Yu, T.F. Quatieri, J.R. Williamson, and J.C. Mundt, "Cognitive Impairment Prediction in the Elderly Based on Vocal Biomarkers," paper in *Proceedings of the 16th Annual Conference of the International Speech Communication Association*, 2015.
46. 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," paper in *Proceedings of the Fourth International Workshop on Audio/Visual Emotion Challenge*, 2013, pp. 65–72.
47. T.F. Quatieri, J.R. Williamson, C.J. Smalt, T. Patel, J. Perricone, D.D. Mehta, et al., "Vocal Biomarkers to Discriminate Cognitive Load in a Working Memory Task," paper in *Proceedings of the 16th Annual Conference of the International Speech Communication Association*, 2015, pp. 2684–2688.
48. F.H. Guenther, *Neural Control of Speech*. Cambridge, Mass.: MIT Press, 2018.

49. K.A. Gola, Tal Shany-Ur, P. Pressman, I. Sulman, E. Galeana, H. Paulsen, et al., "A Neural Network Underlying Intentional Emotional Facial Expression in Neurodegenerative Disease," *NeuroImage: Clinical*, vol. 14, 2017, pp. 672–678.
50. T.F. Quatieri, *Discrete-Time Speech Signal Processing: Principles and Practice*. Upper Saddle River, N.J.: Prentice Hall, 2002.
51. R.L. Rabiner and B.-H.Juang, *Fundamentals of Speech Recognition*. Upper Saddle River, N.J.: Prentice Hall, 1993.
52. P. Ekman and E.L. Rosenberg, eds., *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS),* 2nd ed. New York: Oxford University Press, 2005.
53. T.F. Quatieri and N. Malyska, "Vocal-Source Biomarkers for Depression: A Link to Psychomotor Activity," paper in *Proceedings of the 13th Annual Conference of the International Speech Communication Association*, 2012, pp. 1059–1062.
54. 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," paper in *Proceedings of the Fourth International Workshop on Audio/Visual Emotion Challenge*, 2014, pp. 65–72.
55. B.S Helfer, T.F. Quatieri, J.R. Williamson, L. Keyes, B. Evans, W.N. Greene, et al., "Articulatory Dynamics and Coordination in Classifying Cognitive Change with Preclinical mTBI," paper in *Proceedings of the 15th Annual Conference of the International Speech Communication Association*, 2014.
56. A. Trevino, T.F. Quatieri, and N. Malyska, "Phonologically-Based Biomarkers for Major Depressive Disorder," *EURASIP Journal on Advances in Signal Processing: Special Issue on Emotion and Mental State Recognition from Speech*, vol. 42, no. 1, 2011, pp. 2011–2042.
57. J.R. Williamson, T.F. Quatieri, B.S. Helfer, J. Perricone, S.S. Ghosh, G. Ciccarelli, and D.D. Mehta, "Segment-Dependent Dynamics in Predicting Parkinson's Disease," paper in *Proceedings of the 16th Annual Conference of the International Speech Communication Association*, 2015, pp. 518–522.
58. J.R. Williamson, D.W. Bliss, and D.W. Browne, "Epileptic Seizure Prediction Using the Spatiotemporal Correlation Structure of Intracranial EEG," paper in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 2011, pp. 665–668.
59. J.R. Williamson, D.W. Bliss, D.W. Browne, and J.T. Narayanan, "Seizure Prediction Using EEG Spatiotemporal Correlation Structure," *Epilepsy and Behavior*, vol. 25, no. 2, 2012, pp. 230–238.
60. R. Horwitz, T.F. Quatieri, B.S. Helfer, B. Yu, J.R. Williamson, and J. Mundt, "On the Relative Importance of Vocal Source, System, and Prosody in Human Depression," paper in *Proceedings of the 2013 IEEE International Conference on Body Sensor Networks*, 2013, pp. 1–6.
61. P.J. Jackson and C.H. Shadle, "Performance of the Pitch-Scaled Harmonic Filter and Applications in Speech Analysis," paper in *Proceedings of the 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2000, pp. 1311–1314.
62. P.J. Jackson and C.H. Shadle, "Pitch-Scaled Estimation of Simultaneous Voiced and Turbulence-Noise Components in Speech," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 7, 2011, pp. 713–726.
63. J. Kane, T. Drugman, and C. Gobl, "Improved Automatic Detection of Creak," *Computer Speech and Language*, vol. 27, no. 4, 2013, pp. 1028–1047.
64. T. Drugman, Toolbox, http://tcts.fpms.ac.be/~drugman/Toolbox/.
65. D.A. Reynolds, T.F. Quatieri, and R.B. Dunn, "Speaker Verification using Adapted Gaussian Mixture Models," *Digital Signal Processing*, vol. 10, no. 1, 2000, pp. 19–41.
66. W. Shen, C. White, and T.J. Hazen, "A Comparison of Query-by-Example Methods for Spoken Term Detection," paper in *Proceedings of the 10th Annual IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2009, pp. 2143–2146.
67. B.S. Helfer, J.R. Williamson, B.A. Miller, J. Perricone, and T.F. Quatieri, "Assessing Functional Neural Connectivity as an Indicator of Cognitive Performance," paper presented at the 5th Workshop on Machine Learning and Interpretation in Neuroimaging, Montreal, 11–12 Dec. 2015.
68. American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed. Washington, D.C.: American Psychiatric Association Publishing, 2000.
69. P. Boersma, "Accurate Short-Term Analysis of the Fundamental Frequency and the Harmonics-to-Noise Ratio of a Sampled Sound," paper in *Proceedings of the Institute of Phonetic Sciences*, vol. 17, 1993, pp. 97–110.
70. 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, 1965, pp. 44–49.
71. 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, 1965, pp. 217–224.
72. J.A. Logemann and H.B. Fisher, "Vocal Tract Control in Parkinson's Disease," *Journal of Speech and Hearing Disorders*, vol. 46, no. 4, 1981, pp. 348–352.
73. S. Skodda, "Aspects of Speech Rate and Regularity in Parkinson's Disease," *Journal of Neurological Sciences*, vol. 310, no. 1–2, 2011, pp. 231–236.
74. R. Netsell and D. Lefkowitz, "Speech Production Following Traumatic Brain Injury: Clinical and Research Implications," *ASHA Neurophysiology and Neurogenic Speech and Language Disorders*, vol. 2, no. 4, pp. 1–8.
75. B.S. Helfer, T.F. Quatieri, L. Keyes, B. Evans, W.N. Greene, T. Vian, et al., "Voice and Ocular Dynamics in Classifying Cognitive Change with Preclinical mTBI," poster presented at the Fourth Annual Traumatic Brain Injury Conference, Washington, D.C., 16–17 Apr. 2014.

76. ImPACT Applications Inc., "Immediate Post-Concussion Assessment Testing (ImPACT®) Test Technical Manual Technical Manual and Psychometric Data," available at impacttest.com, 2013.
77. A.C. Lammert, J.R. Williamson, A. Hess, T. Patel, T.F. Quatieri, H.J. Liao, et al., "Noninvasive Estimation of Cognitive Status in Mild Traumatic Brain Injury Using Speech Production and Facial Expression," paper in *Proceedings of the Seventh International Conference on Affective Computing and Intelligent Interaction*, 2017, pp. 105–110.
78. K. Watanabe and S. Funahashi, "Neural Mechanisms of Dual-Task Interference and Cognitive Capacity Limitation in the Prefrontal Cortex," *Nature Neuroscience*, vol. 17, 2014, pp. 601–611.
79. P. Zarjam, J. Epps, and F. Che, "Spectral EEG Features for Evaluating Cognitive Load," paper in *Proceedings of the 33rd Annual International Conference of IEEE Engineering in Medicine and Biology Society*, 2011, pp. 3841–3844.
80. R. Patel, K. Connaghan, D. Franco, E. Edsall, D. Forgit, L. Olsen, et al., "'The Caterpillar': A Novel Reading Passage for Assessment of Motor Speech Disorders," *American Journal of Speech-Language Pathology*, vol. 22, no. 1, 2013, pp. 1–9.
81. D.P. Redmond, M.L. Johnson, D.R. Thorne, G. Belenky, T.J. Balkin, W.F. Storm, et al., "Fatigue Models for Applied Research in Warfighting," *Aviation, Space, and Environmental Medicine*, vol. 75, no. 3, 2004, pp. 44–60.
82. J. Sloboda, A. Lammert, J. Williamson, C. Smalt, D.D. Mehta, I. Curry, et al., "Vocal Biomarkers for Cognitive Performance Estimation in a Working Memory Task," paper in *Proceedings of the 19th Annual Conference of the International Speech Communication Association*, 2018.
83. M.A. Nolan, J.R. Williamson, L.J. Brattain, M.D. Eddy, J.M. Moran, C.J. Smalt, and T.F. Quatieri, "Predicting Variations in Cognitive Load: A Multimodal Approach," poster presented at Neuroscience 2016, San Diego, 12–16 Nov. 2016.
84. T.F. Quatieri, B. Yu, J. Perricone, A. Lammert, M. Nolan, D. Mehta, and J.S. Palmer, "Energy and Correlation Analysis of Gait and EEG Representations During an Auditory Working Memory Task, Movement and Cognition," poster presented at the Movement and Cognition International Conference at Harvard University School of Medicine, Boston, 27–29 Jul. 2018.
85. R. Horwitz-Martin, T. Quatieri, A. Lammert, J. Williamson, Y. Yunusova, E. Godoy, et al., "Relation of Automatically Extracted Formant Trajectories with Intelligibility Loss and Speaking Rate Decline in Amyotrophic Lateral Sclerosis," paper in *Proceedings of the 18th Annual Conference of the International Speech Communication Association*, 2017.
86. B. Yu, J.R. Williamson, J.C. Mundt, and T.F. Quatieri, "Speech-Based Automated Cognitive Impairment Detection from Remotely-Collected Cognitive Test Audio," *IEEE Access*, vol. 6, no. 1, 2018, pp. 40494–40505.
87. T.F. Quatieri, J. O'Rourke, L. Nowinski, D. Hannon, A. Lammert, J. Williamson, et al., "Quantifying Fine Motor Dependencies in Autism Spectrum Disorder," *Neuroscience*, 2018.
88. World Health Organization, "Clinical Management of Severe Acute Respiratory Infection (SARI) When COVID-19 Disease is Suspected: Interim Guidance Report," 13 Mar. 2020, available at https://www.who.int/publications-detail/clinical-management-of-severe-acute-respiratory-infection-when-novel-coronavirus-(ncov)-infection-is-suspected.
89. L. Mao, M. Wang, S. Chen, Q. He, J. Chang, C. Hong, et al., "Neurological Manifestations of Hospitalized Patients with COVID-19 in Wuhan, China: A Retrospective Case Series Study," *JAMA Neurology*, published online 10 Apr. 2020.
90. C.H. Yan, F. Faraji, D.P. Prajapati, C.E. Boone, and A.S. DeConde, "Association of Chemosensory Dysfunction and Covid-19 in Patients Presenting with Influenza-like Symptoms," *International Forum of Allergy & Rhinology*, published online 12 Apr. 2020.

About the Authors

Thomas F. Quatieri, a senior member of the technical staff in the Human Health and Performance Systems Group, is involved in bridging human language and bioengineering research and technologies. Within this group, he has initiated and developed major R&D and technology transition programs in speech and auditory signal processing and neuro-biophysical modeling with application to detection and monitoring of neurological, neurotraumatic, and stress conditions. He has been an author on more than 200 publications, holds 11 patents, and authored the textbook *Discrete-Time Speech Signal Processing: Principles and Practice*. He also holds a faculty appointment in the Harvard-MIT Speech and Hearing Bioscience and Technology Program. He developed the MIT graduate course Digital Speech Processing and is active in advising graduate students on the MIT and Harvard campuses. He is a recipient of four IEEE Transactions Best Paper Awards and the 2010 MIT Lincoln Laboratory Best Paper Award. He led the Lincoln Laboratory team that won the 2013 and 2014 AVEC Depression Challenges and the 2015 MIT Lincoln Laboratory Team Award for their work on vocal and facial biomarkers. He has served on the IEEE Digital Signal Processing Technical Committee, the IEEE Speech and Language Technical Committee, and the IEEE James L. Flanagan Speech and Audio Awards Committee. He has served on National Institutes of Health and National Science Foundation panels, been an associate editor for the *IEEE Transactions on Signal Processing*, and is an associate editor of *Computer, Speech, and Language*. He has been an invited speaker at conferences, workshops, and meetings, most recently as a keynote speaker at the 2019 Speech Science and Technology Conference. He is a Fellow of the IEEE and a member of Tau Beta Pi, Eta Kappa Nu, Sigma Xi, International Speech and Communication Association (ICSA), Society for Neuroscience, Association for Research in Otolaryngology (ARO), and the

Acoustical Society of America (ASA). He holds a bachelor's degree (summa cum laude) from Tufts University, and master's, electrical engineering, and doctoral degrees from MIT.

James R. Williamson is a member of the technical staff in the Human Health and Performance Systems Group. His work focuses on detecting and modeling neurocognitive and physical state through the analysis of speech, facial expressions, gait, and body-sensing modalities. Prior to joining Lincoln Laboratory, he earned a bachelor's degree in psychology from the University of Massachusetts Amherst and a doctoral degree in cognitive and neural systems from Boston University. He also served as a research assistant professor at Boston University.

Adam C. Lammert was a member of the technical staff in the Bioengineering Systems and Technologies Group beginning in 2015, primarily focusing on neurocognitive analysis, motor control modeling, and speech signal processing. He currently holds the position of assistant professor at Worchester Polytechnical Institute. From 2006 to 2008, he was lab manager of speech and hearing research at the Veterans Affairs Northern California Health Care System in Martinez, California. He earned a doctoral degree in computer science from the University of Southern California, where he was a member of the Signal Analysis and Interpretation Laboratory and the Hearing and Communication Neuroscience Program at the House Research Institute. He was awarded the Raymond H. Stetson Scholarship in Phonetics and Speech Science by the Acoustical Society of America in 2013. Prior to joining Lincoln Laboratory, he was a visiting assistant professor in the Computer Science Department at Swarthmore College. He has published more than 45 papers, several of which have been awarded special status. He earned a bachelor's degree in cognitive science from Vassar College and a master's degree in computer science from North Carolina State University.

Kristin Heaton joined the Military Performance Division of the U.S. Army Research Institute of Environmental Medicine in Natick, Massachusetts, as a research psychologist in 2005. In July 2019, she took on the role of deputy chief in the Military Performance Division and currently holds the position of supervisory research psychologist. She earned a doctoral degree in psychology from the University of Maryland, College Park in 2002. Specializing in the field of neuropsychology, she completed her postdoctoral clinical fellowship in neuropsychology with the Boston Consortium in Clinical Psychology, a joint program of the VA Boston Healthcare System, Boston University, and Harvard University. Previously, she was an adjunct research assistant professor at Boston University School of Public Health. Her research focuses broadly on the development, application, and validation of tools and techniques for assessment of neurobehavioral health and performance in military operational and training environments. Her current research portfolio includes studies examining the following: unobtrusive, naturally occurring behavioral markers (e.g., speech, facial expression, and physiological responses) for prediction and monitoring of warfighter performance (cognitive, mental, physical); the impacts of military occupational, environmental, and neurotoxicant (fuel products, pesticides) exposures on warfighter neurological health and performance outcomes; and novel approaches such as electrical brain stimulation for optimizing and enhancing warfighter physical, mental, and cognitive performance. She is currently an active member of several federal agency (Department of Defense) and international (NATO) working groups and committees.

Jeffrey S. Palmer is the leader of the Human Health and Performance Systems Group at Lincoln Laboratory. He has oversight of multiple research programs that focus on health, human performance, objective neurocognitive analytics, and biosensing via wearable, ingestible, and implantable devices. In 2010, he helped to create the first Army War College Fellowship at Lincoln Laboratory and the MIT Security Studies Program. He has given presentations at international conferences and authored book chapters and technical articles on DNA biometrics and forensics, biomechanics, cell biology, materials science, soldier nanotechnology, biological-chemical defense, polymer science, high-energy lasers, microelectronics packaging, wearable biomedical sensing in extreme environments, and neurocognitive technologies. He has served on editorial boards for journals in biomechanics, molecular science, biomedical informatics, and biosensors. He has chaired technical conferences for the National Science Foundation, Department of Homeland Security, and the IEEE. Currently, he is the vice chair (and chair-elect) of the IEEE Engineering in Medicine and Biology Society's (EMBS) Technical Committee on Wearable Biomedical Sensors and Systems and the EMBS conference editorial board for tissue engineering and biomaterials. In addition, he has served as an advisor on senior military studies of enhancing health and performance, and led a multi-agency U.S. government effort to develop automated rapid human DNA analysis capabilities for field biometrics and forensics applications. Prior to working at Lincoln Laboratory, he worked at research laboratories at IBM and GE, and at the Physical Sciences Laboratory at New Mexico State University. He holds a bachelor's degree with a minor in mathematics from New Mexico State University, a master's degree from Rensselaer Polytechnic Institute, and a doctorate with a minor in bioengineering from MIT, all with majors in mechanical engineering.