Received June 12, 2015, accepted July 1, 2015, date of publication July 16, 2015, date of current version August 4, 2015.

Digital Object Identifier 10.1109/ACCESS.2015.2457392

Challenges in Concussion Detection Using Vocal Acoustic Biomarkers

CHRISTIAN POELLABAUER[1], (Senior Member, IEEE), NIKHIL YADAV[1], LOUIS DAUDET[1], SANDRA L. SCHNEIDER[2], CARLOS BUSSO[3], (Senior Member, IEEE), AND PATRICK J. FLYNN[1], (Fellow, IEEE)

[1]Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN 46556, USA
[2]Department of Communicative Sciences and Disorders, Saint Mary's College, Notre Dame, IN 46556, USA
[3]Department of Electrical Engineering, University of Texas at Dallas, Richardson, TX 75080, USA

Corresponding author: C. Poellabauer (cpoellab@nd.edu)

This work was supported in part by the National Science Foundation under Grant IIS-1450349 and in part by the National Football League and General Electric Health through the NFL/GE Head Health Challenge.

ABSTRACT Acoustic metrics extracted from speech have the potential to serve as novel biomarkers for a variety of neurological and neurodevelopmental conditions, as is evidenced by the rapidly growing corpus of research articles studying the links between brain impairments and speech. In this paper, we discuss the advantages and the disadvantages of speech biomarkers and the various challenges in the design and the implementation of portable speech-based diagnostic and assessment tools. Furthermore, we provide a case study, presenting our experiences in developing an assessment tool for the detection of mild traumatic brain injuries (concussions) and discuss the challenges in obtaining and analyzing large sets of speech recordings that can be used to study the impact of brain injuries on vocal features.

INDEX TERMS Speech analysis, vocal features, mild traumatic brain injuries, concussions, speech recognition.

I. INTRODUCTION

Traumatic brain injuries (TBIs) are a disruption of normal brain function due to a bump, blow, or jolt to the head, e.g., caused by car accidents, explosive blasts, and head-to-head hits in contact sports. In the U.S. alone, TBI accounts for an estimated 1.6-3.8 million sports injuries every year [1] and nearly 300,000 concussions are being diagnosed among young athletes every year [2], [3]. Concussions (also referred to as *mild traumatic brain injuries* or *mTBI*) are very common consequences of motor vehicle accidents, falls, and sports injuries. Athletes in sports such as football, hockey, and boxing are at a particularly large risk, e.g., six out of ten NFL athletes have suffered concussions, according to a study conducted by the American Academy of Neurology in 2000. In addition, TBI is also very frequent among soldiers and often called the ''signature wound'' of the Iraq and Afghanistan wars. The potential short- and long-term impacts on the health and well-being of individuals with brain injuries are extensive. For example, individuals with mTBI may display a range of somatic, affective, and cognitive symptoms such as headaches, depression, loss of memory, and loss of brain function. These symptoms are collectively known as Post Concussion Syndrome (PCS) and may persist for weeks or months. The effects of concussions and other brain injuries can be devastating if they remain unrecognized for long durations of time. Recent events in professional sports (such as the suicide of Chicago Bears safety Dave Duerson in 2011) have raised awareness of the many effects of diseases such as chronic traumatic encephalopathy (CTE), which has been tied to depression, dementia, and suicide. A number of studies [4], [5] have shown that head injuries can lead to other long-term health issues such as an elevated incidence of Alzheimer's Disease and a reduced age of onset for Alzheimer's [6]. The work presented in [7] has shown that CTE in boxers leads to dementia developing at a higher rate and a younger age, compared to the general population. Finally, in another recent study [8], it was reported that the diagnosis of mild cognitive impairment (MCI) was more common among football players who reported three or more concussions compared to players reporting none.

The dramatic impact of neurological degenerative pathologies, trauma, stroke, psychiatric disorders, and other disorders affecting the brain on the quality of life and life expectancy is a growing concern. These impacts on the short- and long-term health can be particularly dramatic for young adults whose brains have not yet fully developed.

Unfortunately, it is estimated that almost 90% of concussions remain undetected and therefore untreated [9]. Reasons why so many concussions remain undetected include the fact that brain injuries are difficult to diagnose due to the subtlety of symptoms, a lack of reliable biomarkers that can be used to quickly and non-intrusively detect signs of concussions, and a lack of ''sideline'' assessment tools and technologies that can capture and analyze these biomarkers in near real-time wherever needed. Traditional diagnostic tools used to diagnose brain injuries (such as CT scan, MRI, or x-ray) do not detect concussions because they do not leave any physical traces of damage in the brain. A number of concussion assessment tools are available that can be used to facilitate diagnosis, most of them being neurocognitive tests. One important trend used in sports medicine is the use of *baseline testing* of athletes prior to an injury. These tests are typically administered during the pre-participation physical exam and repeated whenever an athlete might have experienced a concussion. Test results are then compared to the baseline to establish whether a concussion has occurred (i.e., when the scores differ significantly) and to determine the progress of recovery of the athlete during the rehabilitation period. Today, computerized neuropsychological tests are slowly being adopted by schools and sports teams across the country, but the effectiveness and accuracy of existing concussion tests are increasingly being questioned and it is likely that future concussion assessments will rely on a combination of different types of screening tools to increase the accuracy of assessment results [10]. Therefore, there is a need for new concussion biomarkers and screening tools that provide reliable results, but is also easy to use, quick, and can be administered wherever and whenever needed. For example, in athletics, researchers are looking for tools that can be used ''at the sideline'' whenever an athlete may have experienced a concussive hit, without being intrusive and time-consuming. Luckily, the availability of low-cost mobile computing platforms (smartphones and tablets) provides an opportunity for such sideline testing, but only if appropriate biomarkers can be identified.

Over the last couple of decades, a plethora of research results have provided evidence that neurological disorders leave a fingerprint in speech production and speech signal analysis can provide clinical information that can be used to predict certain diseases, diagnose illnesses, and assess disease progression or the effectiveness of treatment regimens [11]–[15]. The primary idea of using speech as biomarker is that brain injuries often manifest themselves by affecting the coordination and timing of the speech motor system, which in turn is reflected in the speech (e.g., distorted vowels, hypernasality, imprecise consonants, etc.) [16]. Speech analysis can be done quickly (even on modern mobile devices such as smartphones and tablets), with minimum inconvenience or intrusion, and with minimal cost (e.g., compared to Nuclear Magnetic Resonance and other advanced technologies). The characteristics of changes in speech and voice caused by neurological diseases, such as Parkinson's disease (PD), amyotrophic lateral sclerosis (ALS), cerebellar demyelination, and stroke, offer an opportunity to provide information for early detection of onset, progression, and severity of these diseases. While previous research results have shown that speech can serve as a biomarker for neurological conditions, very little has been done in trying to coordinate the perceptual features with acoustic measurements to develop an objective, reliable, and accurate diagnostic and assessment instrument. Therefore, this paper first discusses the various challenges encountered in designing and developing diagnostic tools using speech analysis in general and then provides a case study of a speech data collection effort specifically for concussion detection.

II. RELATED WORK

There have been several previous studies related to motor speech disorders and their effects on speech acoustics. Theodoros et al. conducted a study of the speech characteristics of 20 individuals with closed head injuries (CHI) [17]. Their main result was that the CHI subjects were found to be significantly less intelligible than normal (non-neurologically impaired) individuals, and exhibited deficits in the prosodic, resonatory, articulatory, respiratory, and phonatory aspects of speech production. Ziegler and von Cramon discovered an increase in vowel formant frequencies as well as duration of vowel sounds in persons with spastic dysarthria resulting from brain injury [16]. While the focus of our work is on the impacts brain injuries have on speech production, several research efforts have focused on speech processing (i.e., the ability to process and interpret speech). In [18], a variation of the Paced Auditory Serial Addition Task (PASAT) test, which increases the demand on the speech processing ability with each subtest, is used to find out the impact of TBI on both auditory and visual facilities of the test takers. Hinton-Bayre et al. [19] illustrated that tests on speech processing speed were affected by post-acute mTBI on a group of rugby players. Studies have also been conducted on the accommodation phenomenon, where test takers tend to adapt or adjust to unfamiliar speech patterns over time. Research has shown that accommodation is fairly rapid for healthy adults [19]–[21], where the results presented in [21] did not show that subjects with mTBI accommodate slower than healthy subjects. While these prior efforts provide some evidence that neurological conditions have an impact on both speech production and processing, they are typically very limited in the size of the subject pools, focus on only very few specific vocal features, and face other challenges such as poor quality of speech recordings. The purpose of our work is to thoroughly investigate the relationship between brain injury and vocal features that will provide the foundation for the development of novel diagnostic and assessment tools based on speech analysis [16], [17], [22].

Mild forms of traumatic injuries such as concussions have received limited amounts of attention due to the subtlety of the changes in speech compared to the more readily detectable changes caused by many other

neurodegenerative conditions. Parkinson's disease has received the most attention with respect to using speech analysis. PD is a degenerative disease caused by the slow and progressive destruction of neurons and patients characteristically show a variety of symptoms such as tremors, rigidity, and bradykinesia [23], [24]. There are a variety of efforts that have been made attempting to develop early screening systems [12], [13], [25], [26], including techniques based on speech signals [13], [14], [27], [28]. Nearly 90% of individuals with PD develop voice and speech disorders (dysarthria) in the course of their disease [29] and affected patients may complain about a quiet or weak voice and difficulties to initiate speech. The work in [24] has analyzed speech signals to obtain useful clinical information that may be used in order to predict PD. It focuses primarily on the efficient design and implementation of clustering algorithms to quickly identify patients that may require more expensive investigations such as the use of MRI. Another well-known study has been performed by the University of Oxford in collaboration with the National Centre for Voice and Speech, Denver, Colorado [30], where the researchers collected 195 voice recordings with the goal to discriminate healthy people from those with PD.

In addition to PD, a variety of other diseases have received some attention with respect to investigating their relationship to speech. Aphasia is a communication disorder that impacts a person's ability to use and comprehend language [31], [32]. It is a symptom of brain damage that affects approximately one million Americans and is primarily caused by stroke. One out of four stroke survivors experiences some form of language impairment, e.g., when a stroke damages the frontal and parietal lobes in the right hemisphere of the brain. Analysis of voice data can potentially be used to detect subtle signs of stroke, measure the extend of the damage, and assess the rehabilitation progress [33].

Post-traumatic stress disorder, often abbreviated as PTSD [34], is a complex disorder in which a person's memory, emotional responses, intellectual processes, and nervous system have been disrupted by a traumatic experience. The diagnosis is usually made on the basis of the patient's history and the responses to brief interviews. Positron emission tomography (PET) scans of PTSD patients showed that trauma affects the parts of the brain that govern speech and language, therefore speech analysis may provide insights into the presence of PTSD and a person's response to treatment regimens.

Acoustic features of vocalization of autistic or at-risk infant and children have also been analyzed. The Diagnostic and Statistical Manual of Mental Disorders by the American Psychiatric Association [35] characterizes autism as severe and pervasive impairments in the development of reciprocal social interaction, verbal and nonverbal communication skills, and stereotyped patterns of behaviors and interests. It is a pervasive developmental disorder that typically manifests itself in the first three years of life [36]. Early identification of ASD (autism spectrum disorder) has become a national priority as an increasing number of studies provide evidence that the impairments associated with ASD can be ameliorated through intensive early, targeted, autism-specific services [37], [38]. For example, in several papers [36], [39]–[47], researchers present the analysis results of cry signals of infants, showing that certain acoustic metrics such as pitch and formant frequencies can vary between healthy children and children later diagnosed with ASD.

As part of its investigation of the EXXON VALDEZ accident and oil spill, the National Transportation Safety Board (NTSB) examined the captain's speech for alcohol-related effects, with speech samples obtained from marine radio communications tapes [48]. The speech samples were tested for slowed speech, speech errors, mis-articulation of difficult sounds (''slurring''), and audible changes in speech quality. It was found that speech immediately before and after the accident displayed large changes of the sort associated with alcohol consumption. As a consequence, it appears that speech analysis may be a useful technique to provide secondary evidence of alcohol impairment [49]–[51].

Another group of health conditions that has the potential for assessment and diagnosis based on speech analysis is the group of mental disorders. The Global Burden of Disease Study by the World Health Organization has found that mental health difficulties are currently the leading cause of disability in developed countries [52] and projections indicate that the global burden of mental health difficulties will continue to rise in the coming decades. In the UK, mental health has overtaken unemployment as the nation's most expensive social problem [53]. As an example, there are no established objective biomarkers for schizophrenia and it has been previously reported that there are notable qualitative differences in the speech of schizophrenics. Vowel production of people with schizophrenia has been analyzed in [55] and [56] to determine whether a quantitative acoustic and temporal analysis of speech may be a potential biomarker for schizophrenia.

Finally, there has also been prior work studying biomarkers for multiple diseases simultaneously, e.g., the work in [56] presents a rationale for acoustic analysis of voices of neurologically diseased patients, providing preliminary data from patients with myotonic dystrophy, Huntington's disease, Parkinson's disease, and amyotrophic lateral sclerosis, as well as from individuals at risk for Huntington's disease. This work showed that noninvasive acoustic analysis may be of clinical value for early diagnosis and for documenting progression for various diseases.

III. MOTIVATION AND BACKGROUND

A. MOTOR SPEECH DISORDERS

Under normal circumstances, speech is produced with ease and without thought to the complexity of the underlying mechanisms that are employed. But in fact the act of speaking requires the integration of numerous neuromuscular, neurocognitive, and musculoskeletal activities.

TABLE 1. Types of dysarthria.

Type of Dysarthria	Perceptual and Acoustic Features	Localization
Flaccid	Breathiness, Weakness, Hypotonia, Reduced reflexes, Nasal emission, Fasciculations/fibrillations, Articulatory breakdowns	Lower motor neuron lesion
Spastic	Harshness, Slow rate, Hypertonia, Effortful speech, Excessive stress, Hypernasality, Pitch breaks, Short phrases	Multiple upper motor neuron lesions
Hypokinetic	Hypernasality, Rapid/"blurred" alternation motion rates (AMRs), Breathiness, Reduced stress, Inappropriate silences, Short rushes of speech, Tremor	Basal ganglia control circuit
Hyperkinetic	Hypernasality, Inappropriate involuntary movement, Unequal stress/breaks, Harshness, Voice tremor, Slow/fast rate	Basal ganglia control circuit
Ataxic	Imprecise/irregular articulation, Irregular AMRs, Excessive/unequal stress, Prolonged phonemes, Prolonged intervals, Slow rate	Cerebellum
Unilateral Upper Motor Neuron	Breathiness, Slow rate, Weakness, Articulation breakdowns, Incoordination	Unilateral upper motor neuron lesion
Mixed	Any combination of the above features	More than one location

An average speaking rate, which consists of producing 14 sounds (phonemes) per second, requires the execution of 100 different muscles containing on average 100 motor units generating 140,000 neuromuscular events [57]. This motor speech execution is proceeded by a planning/programming phase that involves translating abstract linguistic-phonological representation into a code that is then used by the motor system to generate those specific moves. Since the motor speech system is a dynamic and complex act involving many systems and connections, disruptions can occur at any or all levels of planning and execution. These breakdowns result in predictable patterns of disturbances [31], [32]. Motor speech disorders are a group of disorders resulting from disturbances in the central or peripheral nervous systems that affect the motor programming/programming and execution/control/regulation of the motor movements.

Motor speech disorders can be categorized into two distinctive types. Apraxia of speech is an impairment in the ability to plan and program the movements of speech, usually due to the result of a cortical lesion in the dominant hemisphere. Dysarthria is a disorder involving the execution of movement that affects the strength, speed, range, steadiness, tone, or accuracy of movement that can affect respiration, resonance, phonation, and articulation [31], [32]. Dysarthria is further categorized into different types associated with specific types of neurological deficits and disturbances, each distinguishable by specific auditory perceptual and acoustic features (see Table 1). Thus, any abnormality of the motor speech system can shed light on the integrity of the neurological system. Currently, the most widely used method of determining disturbances in the motor speech system is through perceptual detection, which lends itself to subjective bias. While computerized methods are available, they lack convenience (i.e., bulky computer based equipment), training in use of the equipment is complex, they are expensive, and foremost, there is a lack of evidence to support the contribution of the instrumentation in diagnosis and treatment of motor speech disorders. However the need for accurate, quick, objective measures is sorely needed to assist in the diagnosis of the motor speech disorder. This then will lead to correct and more cost-effective treatment, as well as help with the understanding of the underlying pathophysiological causes associated with concussion and brain injuries in general.

B. SPEECH AS A BIOMARKER

Our research focuses on the identification of new biomarkers that can be used to detect early (and often very subtle) signs of new or deteriorating neurological conditions and traumatic brain injuries, while being easy-to-use and minimally invasive. Over the last couple of decades, several research efforts have provided evidence that neurological disorders leave a fingerprint in voice and speech production [31], [32]. Speech signal analysis can provide clinical information that can be used to predict certain diseases, provide information about the neurological location of specific diseases, and assess disease progression or the effectiveness of treatment regimens [11]–[15], [24], [48], [56], [58]. There are primarily two methods (acoustic and perceptual) to assess the motor speech characteristics associated with neurological disease diagnosis. The most common method used is the auditory-perceptual method (which involves a professional hearing and seeing the motor speech changes associated with neurological disordered processes) for clinical assessment, judgements, and decisions regarding functional change. Obviously this method is subjective, not reliably quantifiable, and can be influenced by clinician bias. In addition, sometimes these changes in motor speech are so subtle that they typically cannot be recognized by perceptual features alone. The need is for an integration of the auditory-perceptual method with a more objective non-biased measurement of the disordered speech features to accurately extract the most relevant acoustic features in both time and frequency domain. The primary hypothesis of our work is that brain injuries manifest themselves by affecting the motor speech system, which in turn is reflected in vocal feature changes (e.g., rate, voice quality, loudness, resonance, vowel distortions, hypernasality, imprecise consonants, etc.), as evidenced by prior research such as the work presented in [16] and [23]. Toward this end, thorough and large-scale data collections and investigations are required to lead to a better understanding of the relationship of mTBI and speech,

TABLE 2. Speech protocol.

Test	Text Displayed on Screen	Display Duration	Description
1	Application, Participate, Education, Difficulty, Congratulations, Possibility, Mathematical, Opportunity	1.5 seconds per word	Participant reads out multisyllabic words. Test designed to study articulation, stress, and prosody
2	PUT the book here put the BOOK here put the book HERE	10 seconds	Participant reads and stresses different parts of the sentence, i.e., put, book, here; test designed to capture intonation stimulability
3	We saw several wild animals	5 seconds	Participant reads simple sentence to test standard syllabic rate
4	pa	5 seconds	Participants repeat the pa sound as quickly as possible. Alternating motion rate is captured
5	ka	5 seconds	Participants repeat the ka sound as quickly as possible. Alternating motion rate is captured
6	pa-ta-ka	5 seconds	Participants repeat the pa-ta-ka sound as quickly as possible. Sequential motion rate is captured
7	Aahhhh	5 seconds	Voice quality and tremor captured as participants are asked to sustain the sound

such that a reliable speech-based mTBI test can be developed. This paper reports our challenges from building a portable speech data collection app, to collecting large-scale speech data on youth athletes, to the speech processing and analysis to extract a series of identifiable and relevant speech features for concussion detection. The hope is that these insights, experiences, and outcomes will pave the way to future assessment tools that are fast, objective, easy-to-use, and accurate, and provide a more reliable basis for making return-to-game decisions after concussions have been confirmed.

IV. CHALLENGES IN THE DESIGN AND IMPLEMENTATION OF A SPEECH-BASED DATA COLLECTION TOOL

This section describes the various challenges in designing and implementing speech-based tools for both research purposes and real-time diagnostics and assessment, focusing on challenges in application design, speech processing, noise management, user interface design, and data analysis, among others.

A. SPEECH PROTOCOL DESIGN

As discussed in Section III-A, speech production is a complex and integrated system that requires the involvement of about 100 different muscles, with different sounds requiring the involvement of different sets of muscles performing in different ways. An example of this is the difference between front, middle, and back consonant and vowel sounds. While front consonant sounds such as *t* and *d* require involvement of the muscles in the front of the speaker's mouth (e.g., lips, teeth, front of the tongue), back sounds, such as *g* and *k*, rely more on the soft palate and back of the tongue. Words then combine these sounds, often in complex ways, e.g., a word such as *crisp* "moves" from the back of the mouth through the middle of the mouth to the front while articulated. Most prior efforts in speech analysis for health purposes relied on "random" vocalizations, e.g., speech extracted from opportunistically obtained recordings or video/audio captures. In these cases, the quality of the acoustic features and the dependence of speech on brain impairments may vary significantly from recording to recording and from subject to subject. However, if a specific *speech protocol* is used, i.e., the subject is requested to articulate very specific sounds, words, or phrases, the quality is controllable, and it is easier to compare acoustic features across different subjects. This requires a very careful design of the speech protocol, controlling for: the quality of the production, the ease of administration, i.e., low cognitive/language load, and omitting any words or phrases that might trigger emotional responses. For example, initial versions of our speech protocol included a series of words that begin with the letter *h*. One of these words was *hell* and a surprisingly large number of youth athletes displayed audible changes in tone, hesitation, or even outright refusal to speak the word into the microphone.

The current version of our speech protocol is shown in Table 2; the protocol consists of seven brief tests, each with a different type of text shown on the screen, where the subject reads the text into the computer's microphone as instructed by the test application. The third column in the table shows the test duration, i.e., how long the text is displayed, followed by a brief description of the test in the fourth column. The tests were designed to capture the salient features that are influential to speech production in general and can directly capture the deviant motor speech characteristics that might be associated with concussion. These features include speed, strength, range, accuracy, and steadiness of movement. These features interact and impact each other such that if strength of the motor components of speech are affected, speed, tone, range of movement, accuracy, and steadiness might also be affected. The details of each test are described below.

- *Test 1:* A series of multisyllabic words are shown on the device's screen for exactly 1.5 seconds each. It is anticipated that subjects with mTBI would find it more difficult to correctly enunciate these words. The words consist of 4-5 syllables, varying from front consonant [p, b, t, d, f, m, θ] and front vowel [i, I, e, æ] sounds, to middle consonant [l, s, z, n, ʃ, j, r], middle vowel [ʌ] sounds, to back consonant [k, g, h], back vowel [u, o, a] sounds. During normal speech production, movements are rapid and produced without effort. Any difficulty noted during this task would indicate deficits in range, accuracy, and speed of movement.

- *Test 2:* In the second test, the sentence *put the book here* is displayed on the screen three times; first with emphasis on *put*, then *book*, and then *here*. This test examines the persons ability to put stress on different words during running speech. Abnormalities in strength and range during normal speech production can influence the prosodic features of speech, resulting in the restricted ability to produce stress.
- *Test 3:* The third test requires the participant to read the sentence *we saw several wild animals*. The standard syllabic rate may be affected and cause perceptual differences in articulation. This test also examines the accuracy, strength, and speed of muscle movement. Accuracy is the result of coordinated and well-timed movement.
- *Tests 4, 5, and 6:* In these tests, the subject is asked to repeat the syllables *pa*, *ka*, and *pa-ta-ka* as quickly as possible to measure alternating and sequential motion rates (AMR and SMR, respectively) or diadochokinetic rates. These are used to determine the speed, accuracy, and regularity of motor movement. Most individuals can produce these syllables 5-7 times/second. The test using the syllables *pa-ta-ka* measures the ability to produce the syllables rapidly and in proper sequence.
- *Test 7:* Subjects attempt to sustain the *Aahhhh* sound for at least 5 seconds. Sustaining this vowel sound allows us to examine muscle tone and steadiness of the tone. Alterations in tone can occur during all components of speech production and can affect perceptual and acoustic speech measures. Usually, with normal speech production, there are no interruptions or oscillations in vocal tone. Most individuals will automatically produce the tone at their habitual pitch and loudness level for approximately 8-9 seconds. If unsteadiness of tone is heard, this can be the sign of a neurological disorder such as a concussion.

In summary, the more complex the words or utterances, the higher the likelihood of errors and variations in the speaking of these words. At the same time, the selected words and phrases have been chosen, because they require involvement of different parts of the mouth, tongue, and soft palate in the back of the throat as they contain two or more very different syllables. The examination of the efficiency and accuracy of the motor speech system can provide information about the integrity of the neurological system. Note that different tests provide different opportunities for speech analysis, e.g., single words cannot be used to measure AMR and SMR, but they can be used to measure timing characteristics, stress, and frequency composition of the speech. On the other hand, continuous speech can be used to measure speaking rate or variations in speed, intensity, etc., which could be difficult to measure for short words. While individual words are rather straightforward to detect and analyze using standard speech recognition tools, analysis of continuous speech can be more challenging, e.g., to accurately match the spoken sounds to the sounds expected by the test. This can be particularly challenging for advanced neurological conditions, where mispronunciations or skipped syllables may be common (Section IV-C discusses the challenges in speech processing).

Another important choice in the design of the speech protocol is how the vocal features are evaluated and there are two primary approaches. First, an important trend used in sports medicine is the use of *baseline testing* of athletes prior to a potential injury. These tests are typically administered during a *pre-participation physical exam* and repeated whenever an athlete might have experienced a concussion and the scores from the baseline are compared to the later scores. In speech-based assessment, this is different in that there are no ''scores'' obtained from the tests; instead, each recording yields a series of acoustic features and we form the difference between the features obtained from a subject's recording and that same subject's baseline. For example, the *pa-ta-ka* test described above results in a sequential motion rate and the change in this rate in comparison to the rate measured in the baseline recording is then used for analysis (e.g., a significant reduction in this rate could indicate a neurological problem). A challenge in this approach is then to identify which acoustic features will vary more than others as a consequence of a brain injury (thereby showing greater promise as a potential biomarker) and also how to quantify these changes. This baseline-based technique is the primary approach utilized in our work. In contrast, vocal features can also be compared to a *speech norm*, i.e., vocal features of a typical healthy person. A main advantage of this approach is that no prior baseline of a subject is required. While baselining is typically no problem in sports, access to baseline recordings may be difficult for older adults undergoing testing for neurodegenerative diseases or analyzing the speech of a soldier in a remote area. However, there are numerous disadvantages to norm-based assessment, e.g., these norms must be established through analysis of sufficiently large sets of recordings, ideally divided into groups of subjects of different ages, genders, and other medical conditions. For example, many of the vocal features of a 60 year old healthy female will differ significantly from a male adolescent with a neurodevelopmental condition. Accents and dialects are also very likely to impact feature comparison. Therefore, a baseline-based approach will typically be the preferred approach unless access to baseline recordings is not possible. If a norm-based approach is used, the feature comparison should occur with a norm that matches the subject's age, gender, and various potential confounders as much as possible.

B. NOISE MANAGEMENT

The accuracy of speech analysis (recognition, word onset detection, vocal feature extraction, etc.) degrades severely when speech is recorded in adverse acoustical environments. In recent years, significant advances have been made in the area of developing robust speech processing systems that are able to provide satisfactory results even in the presence of environmental noise, including the use of microphone

arrays [59] and filtering techniques [60]. When building a speech-based diagnostic tool, our primary concerns are (1) inaccuracies in the detection of the boundaries of words and phonemes, (2) the inability to extract the desired vocal features, and (3) significant errors in the vocal features. However, while the primary goal of automated speech recognition (ASR) systems is to accurately match a recorded word to a dictionary word, given that we have a concrete speech protocol, recognition of the uttered words is not necessarily required. In order to obtain high quality voice recordings, we address the noise challenge by using a high-quality noise-canceling microphone and a continuous (real-time) evaluation of the quality of the recorded speech. While a carefully selected microphone will be able to eliminate most of the potential interferences, a software-based signal quality analysis can be used to determine whether speech quality is acceptable or a test should be retaken.

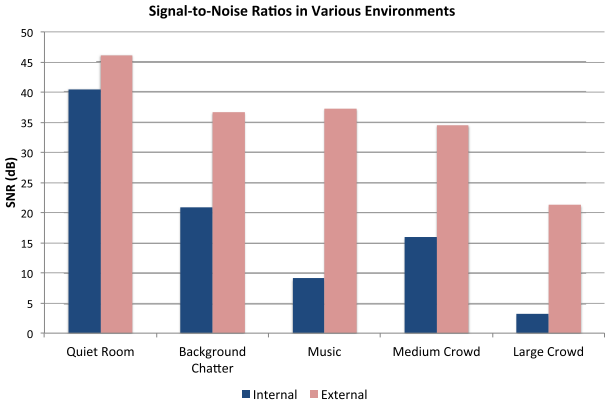

FIGURE 1. Comparison of internal iPad microphone and unidirectional external microphone in various noise scenarios.

In speech analysis, the Signal-to-Noise Ratio (SNR) is a frequently used indicator of signal quality and noise level, and estimated as:

$$10\ log_{10} \frac{\hat{P}_{speech}}{\overline{P}_{noise}} \quad (1)$$

where $\hat{P}_{speech}$ is the speech signal peak power and $\overline{P}_{noise}$ is the mean noise power. Assuming that speech recordings will be obtained using mobile devices such as smartphones or tablets, it is important to understand the limitations of the omnidirectional microphones built into mobile devices. Toward this end, Figure 1 compares the measured average signal-to-noise ratio for the internal microphone of an iPad mini tablet (left bars in the graph) and an external microphone attached to the iPad, specifically, the SHURE SM-10 noise cancellation microphone[1] (right bars), which is a low-impedance, unidirectional dynamic microphone designed for close-talk head-worn applications such as remote-site sports broadcasting and corporate intercom systems. The microphone gain on the iPad was set to the maximum (the gain setting can be varied from 0.1 to 1.0, with 0.6 being the default setting). Figure 1 compares the measured SNR for both microphones in five settings with different quantities and types of noise: an empty room (the only ambient noises were humming sounds from electronic devices and the AC), a room with several people talking about 20-30 feet away from the microphone, a room with music playing from a radio, a cafeteria with mid-morning crowds and activities, and the same cafeteria during the busiest time of the day (lunch hour). In all scenarios, the external microphone outperforms the internal microphone, but the impact on SNR is most dramatic in the noisiest environments. While speech assessments in clinical settings are likely to be performed in relatively quiet and noise-free environments, assessment at the sidelines of sport events will experience much more significant noise pollution, where high-quality microphones are required. However, the results in Figure 1 indicate that most noise issues can be prevented through careful selection of the microphone. But the results also show that even with a high quality microphone, the SNR can drop by more than half in noisy environments, therefore, we also utilize real-time evaluation of SNR on the recording device to determine whether a recording is of satisfactory quality or not. Toward this end, we continuously compute the SNR value of speech while it is being recorded and indicate to the user at the end of the test if the test was accepted or if it should be retaken (in the latter case, the user will be told to either move the microphone closer to the mouth or to move to a quieter location). Given an SNR estimate of a speech recording, we then use the approach proposed in [61] to determine a threshold of 28dB, i.e., if on average the SNR level of a recording is below that limit, a retest is required.

Since signal and noise levels are combined into our speech recordings, we can only estimate the SNR values using prior knowledge about the behavior of the signal and the noise. For example, SNR can be estimated using the technique suggested by NIST,[2] where energy levels are determined over sections of the recording to characterize speech levels and noise levels. Toward this end, a signal energy histogram is generated by computing the root mean squared (RMS) power (in decibels) over a 20 ms window, then the appropriate histogram bin is updated, and finally the window is shifted by 10 ms to repeat this process. This process also includes an implementation of the "direct search" algorithm [62], which can be computationally expensive to run on resource-constrained mobile devices. By default, the algorithm iterates 250 times to obtain a reliable estimate for the SNR value. However, this may be more resource-consuming than necessary, e.g., Figure 2 compares the execution times of the algorithm to the computed SNR values over varying numbers of iterations. At the left of the graph, the default 250 iterations lead to an SNR estimate of about 35dB (red line), but it takes more than 5 seconds to compute this value (blue line).

[1] http://www.shure.com/americas/products/microphones/sm/sm10a-headworn-microphone

[2] http://labrosa.ee.columbia.edu/~dpwe/tmp/nist/doc/stnr.txt

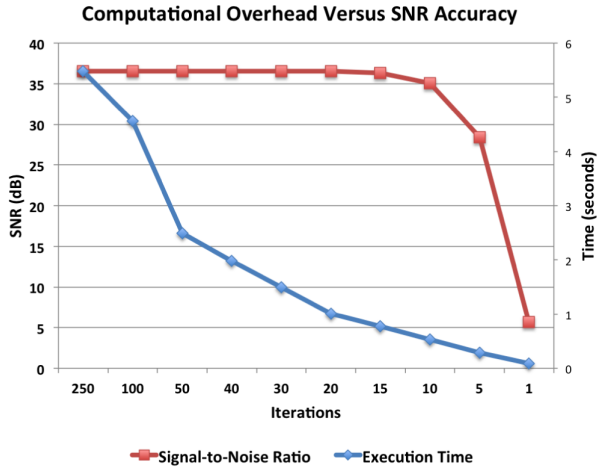

FIGURE 2. Comparison of SNR accuracy and computational overheads of SNR estimation technique.

When we reduce the number of iterations, we can see that the execution overheads decrease quickly, while the SNR estimate remains unchanged until about 10 iterations. That means that it is possible to significantly reduce the number of iterations, without impacting the accuracy, but with a drastic reduction in overheads (e.g., for 10 iterations, the overhead reduces to less than one second).

C. SPEECH PROCESSING

Before we can measure vocal features in speech, it is necessary to first process the audio data, e.g., to perform phonetic transcription (e.g., vowel boundary, syllable rate, etc.). To that end, we rely on current speech processing techniques to process the audio before the analysis. ASR is the process of converting a speech signal into its corresponding sequence of words or other linguistic entities [63], [64]. Specifically, we use the PocketSphinx tool [65], which is one of the few currently available speech recognition toolkits based on Hidden Markov Models (HMMs) and compatible with portable devices.

If W is the word sequence and O is the feature vector, we can estimate $P(W|O)$, i.e., the most likely word sequence given the observations, using the following equation:

$$\underset{W}{arg\ max}\ P(W|O) \propto P(O|W)P(W), \tag{2}$$

where $P(O|W)$ is the acoustic model implemented with HMMs [66]. The structure of the HMMs is a left-to-right topology with three states per phoneme. We use Gaussian mixture models (GMM) for the observation probabilities. The acoustic features are a 39 dimensional vector with 13 Mel Frequency Cepstral Coefficients (MFCCs) plus their delta and delta-delta features. We use 64 mixtures for the GMM. This configuration is commonly used in related ASR studies. The acoustic models were trained with two large speech corpora with read and spontaneous speech: the Wall Street Journal-based Continuous Speech Recognition Corpus Phase II (WSJ) [67] and the 1996 English Broadcast News Speech (HUB4) [68]. We adapt these HMMs using several of our own recordings using the speech protocol described in Section IV-A to minimize mismatches between train and test recordings. We use the Maximum A Posteriori (MAP) adaptation [69] and the Maximum Likelihood Linear Regression (MLLR) adaptation [70] with 6512 files containing sequences of words. The Word Error Rate (WER) of the adapted ASR was about 9.8% when evaluated with the word sequence recordings.

$P(W)$ in Equation 2 is the language model. Given the speech protocol described in Table 2, we designed a test-dependent language model. For the spoken sequences of words and sentences, the language model is constrained to the prompted entries. This approach improves the phonetic boundary detection. For the AMR and SMR tests (‘‘pa’’, ‘‘ka’’, and ‘‘pa-ta-ka’’), we implemented a grammar with a finite state machine that allows multiple repetitions of the target syllable. For Test 7 (extended vowel), we constrained the grammar to multiple repetitions of the given vowel. Initially, we considered to adapt the acoustic models with samples of the extended vowels, aiming to capture their differences in duration with respect to regular vowels. However, we realized that using the acoustic models with this constrained grammar provided very good accuracy in phone detection and segmentation.

The most challenging task for the language model is Test 6 (‘‘pa-ta-ka’’). The task consists of repeating the sequence of syllables as fast and often as possible, so subjects make many mistakes altering the order of the syllables (e.g.,‘‘pa-ka-ta’’), skipping syllables (e.g.,‘‘pa-ta’’), producing mispronunciations (e.g., ‘‘pa-da-ka’’), replacing syllables (e.g., ‘‘pa-ti-ka’’), and restarting the sequence (e.g., ‘‘pa-pa-ta-ka’’). To derive an effective grammar for this task, transcribers annotated the phonetic content and syllable boundaries of 31 examples from Test 6. We created a finite state machine grammar that captures the common errors made by the subjects. These variations were then properly weighted according to their frequency in this reduced set.

D. ACOUSTIC FEATURES

Once an audio recording has been processed, the next step is to extract various acoustic features that will form the basis of the health diagnostics tool. Speech can be characterized by a variety of different features and these features can be determined for various linguistic elements such as phonemes, vowels, words, or even entire sentences. Previous research has studied various types of features of speech, although some have received significantly more attention than others, such as the fundamental frequency (F0 contour) and formant frequencies F1, F2, etc. Figure 3 shows the time-domain presentation of a brief voice recording (top) and the corresponding spectrogram, i.e., the time-frequency-amplitude presentation of the signal (bottom), as visualized by the

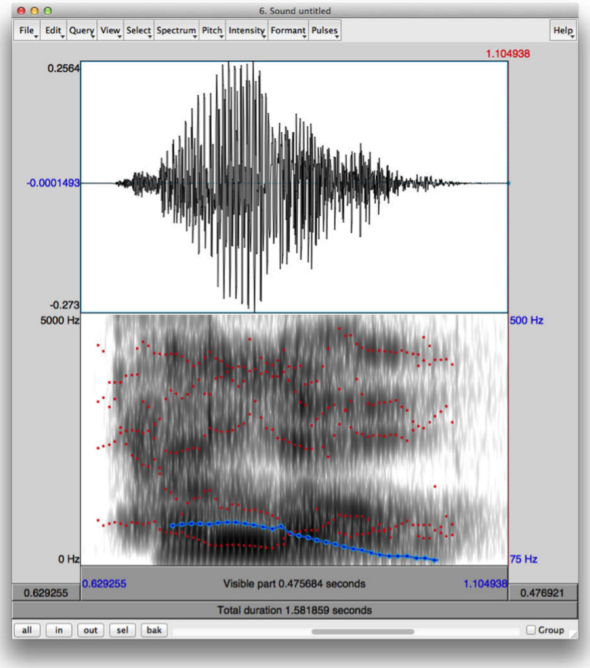

FIGURE 3. Time-domain presentation and spectrogram of a recording of a single word.

Praat software.[3] The identifiable repeating patterns found in speech are called a cycle and the duration of a cycle (glottal pulse) is the pitch period length and F0 is then the inverse of the pitch period length. F0 is shown as the blue dots/lines at the bottom of the spectrogram in Figure 3. Formants are the maxima of the spectral envelope, and are primarily determined by the shape of the vocal tract. These concentrations are usually identified as F1, F2, etc., going from low to high frequencies (shown as the red dots/lines in the spectrogram in Figure 3). The work in [71], one of the most widely cited experiments on the acoustics and perception of vowels, was conducted at Bell Telephone Laboratories by Peterson and Barney in 1952[4] and is an example of a study focusing on pitch and formants. In this study, ten vowels were collected from 76 subjects and fundamental frequency (F0), formant frequencies (F1-F3), and formant amplitudes were extracted and analyzed. Another well known effort for data collection and vowel analysis is the Hillenbrand data set [72], which extends the Peterson and Barney study and uses 139 subjects and 12 vowels, and which also considers formant frequency F4.

While we also consider F0 contour and formants for our analysis, they are only a few examples of acoustic features that we extract and analyze. Table 3 summarizes the 38 acoustic features that are extracted from our speech samples, typically using open-source software solutions such as PocketSphinx, which quite readily extract these features for a given recording. The table provides a brief description for each feature and also indicates for which of the 7 test categories it is expected to be most useful. For example, time measurements such as average and standard deviation of the duration of a linguistic entity, duration of pauses, etc., can be measured for all tests. The diadochokinetic (DDK) rate measures how quickly a person can accurately repeat a series of rapid, alternating phonetic sounds, such as the ''pa'', ''ka'', or ''pa-ta-ka'' sounds in our test protocol, and therefore is not meaningful for the categories of our protocol that consist of single words only. Another feature, jitter, expresses cycle-to-cycle variations of the fundamental frequency. While it is typically measured over long sustained vowel sounds (as in our test category 7), recently the use of jitter over short-term time intervals has also shown promise in analyzing pathological speech [63]. While this list is more comprehensive than most existing studies on variations of acoustic features due to brain abnormalities, it is also a list that is far from complete. That is, our focus has been on identifying what we believe to be some of the more promising features, based on literature or our own work, while ignoring ones that have shown to be less relevant. Furthermore, many acoustic features can be measured or interpreted in various ways, e.g., the F0 contour or formant frequencies shown in Figure 3 are time-variant parameters. In order to compare these frequencies between two recordings (baseline and after a traumatic event), they can be interpreted in different ways: averaged over the entire linguistic entity or over smaller segments, the peak values, the range, the slope, etc. Therefore, identifying and evaluating new acoustic features and new representations or variants of existing features remain open research challenges.

V. CASE STUDY

This section combines the techniques discussed in the previous section and presents a case study of a data collection effort with the purpose of providing a better understanding of the links between brain injuries and speech (i.e., the primary purpose of the application is to collect speech recordings for research purposes and not to provide real-time concussion diagnostics), but also to highlight some of the challenges experienced in performing such a data collection study.

A. DATA COLLECTION APPLICATION

Speech capture is performed on an Apple iPad mini, coupled with the SHURE SM-10 microphone as shown in Figure 4. All speech recordings were collected using the SHURE SM10A external microphone at 44.1kHz, 16 bit, mono, but every recording was immediately downsampled to 26kHz, which is the recommended sampling rate for clinical and empirical voice analysis [73]. In addition to the speech capture, the mobile app provides the subject with an opportunity to provide additional health context. Specifically, the subject provides the following information:

- *Type of Test:* Speech recordings are treated or stored differently depending on the type of test:
  - *Baseline at Rest:* This is typically performed during the pre-participation physical exam,

[3]http://www.fon.hum.uva.nl/praat/
[4]http://www.laps.ufpa.br/aldebaro/repository/pbvowel.htm

TABLE 3. List of acoustic features.

Test	Feature	Acoustic Metric	Description
test1	Time	Average Duration	Average duration of words spoken in test
test1	Time	Standard Deviation Duration	Standard deviation in the words spoken in the test
test2-PUT	Time	Stressed Word Duration	Time taken to say the stressed word "PUT"
test2-PUT	Time	Stress Pause	Pause time before saying the stressed word
test2-PUT	Pitch	F0 Movement	Fundamental frequency movement
test2-PUT	Pitch	F0 Rate	Fundamental frequency rate
test2-PUT	Amp	Intensity Deviation	Deviation in energy intensity (amplitude)
test2-BOOK	Time	Stressed Word Duration	Time taken to say the stressed word "BOOK"
test2-BOOK	Time	Stress Pause	Pause time before saying the stressed word
test2-BOOK	Pitch	F0 Movement	Fundamental frequency movement
test2-BOOK	Pitch	F0 Rate	Fundamental frequency rate
test2-BOOK	Amp	Intensity Deviation	Deviation in energy intensity
test2-HERE	Time	Stressed Word Duration	Time taken to say the stressed word "HERE"
test2-HERE	Time	Stress Pause	Pause time before saying the stressed word
test2-HERE	Pitch	F0 Movement	Fundamental frequency movement
test2-HERE	Pitch	F0 Rate	Fundamental frequency rate
test2-HERE	Amp	Intensity Deviation	Deviation in energy intensity
test3	Time	Average Syllable Duration	This is the syllable duration for the passage (in ms). Many dysarthric speakers have slower rates of speech and the duration increases. This parameter is the inverse of the standard syllabic rate.
test3	Time	Average Pause Duration	This is the pause duration for the passage (in ms). This passage should have no pauses. Therefore, any significant pause time is a variation from normal speech patterns.
test3	Time	Average Diadochokinetic Rate Period	Average DDK period of the subject during this vocalization (ms). The average period is the average time between the consonant-vowel (C-V, e.g., "pa") vocalizations. The period is inversely related to the rate.
test4/test5	Time	Average DDK Rate	The average DDK rate is the number of the C-V (i.e., "pa") vocalizations per second. The rate is inversely related to the average period. Many motor disordered speakers show reduced DDK rates due to decreased articulatory motility.
test4/test5	Time	Standard Deviation in DDK Period	This is the standard deviation of the DDK period (ms). A normal speaker can maintain periodic repetitions while many disordered voices show more variability in their repetition rate, resulting in increased standard deviation of DDK period.
test4/test5	Time	Coefficient of Variation in DDK Period	Degree of rate variation in the period (%). If the C-V vocalization is repeated with little variation in rate, then this number is very small. However, as a speaker varies the rate of DDK over the analysis window, this number increases.
test4/test5	Amp	Standard Deviation in DDK Peak Intensity	Standard deviation of the DDK peak intensity (dB)
test4/test5	Amp	Coefficient of Variation of DDK Peak Intensity	Degree of intensity variation in the peak of each C-V vocalization
test6	Time	Average DDK Period	Same as test 4 and test 5
test6	Time	Average DDK Rate	Same as test 4 and test 5
test6	Time	Standard Deviation of DDK Period	Same as test 4 and test 5
test6	Time	Coefficient of Variation of DDK Period	Same as test 4 and test 5
test6	Amp	Standard Deviation in DDK Peak Intensity	Same as test 4 and test 5
test6	Amp	Coefficient of Variation of DDK Peak Intensity	Same as test 4 and test 5
test7	Pitch	Average F0	Mean of the fundamental frequency
test7	Pitch	Standard Deviation F0	Measure of variability in the data
test7	Pitch	F0 Range	This is a measure of the difference between the maximum and minimum pitch values (in Hz) in the active window (time frame saying the "aahhhh" sound).
test7	Pitch	vF0 (Coefficient of Variation)	The vF0 is defined as the standard deviation F0 divided by the arithmetic mean.
test7	Amp	Standard Deviation F0	Energy measure
test7	Amp	vF0 (Coefficient of Variation)	Coefficient of variation related to energy measure
test7	Pitch	Jitter	The relative average perturbation provides an evaluation of the variability of the pitch period within the analyzed voice sample at a smoothing factor of 3 periods.

where the health of an athlete is evaluated before sports activities begin. Most existing concussion assessment tools perform first testing during that event to establish a baseline for comparison with future test results. In our data collection app, acoustic features are extracted and stored in a database (either locally on the mobile device or, as in our case, on a remote server). Since many subjects perform the test for the first time, the entire test is taken twice to make the subject familiar with the specifics of the test. The first recording is discarded and only the features from the second test are retained.

- *Recording Without Contact:* This data collection is performed after a baseline recording has been obtained and when there is no specific reason to suspect a concussion, e.g., to provide a non-concussed control recording for research purposes.
- *Recording With Contact:* This data collection is performed after a baseline recording has been obtained and when there is reason to believe that a subject

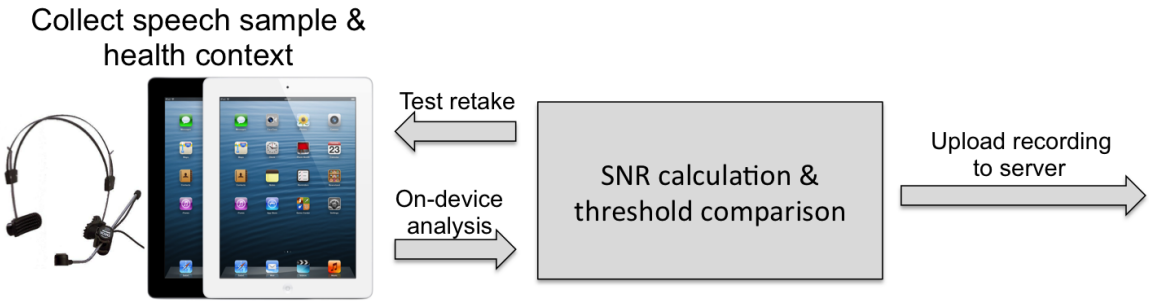

FIGURE 4. Data collection application structure on mobile device.

may have a concussion (e.g., after a severe hit or jolt to the head during a football game). To put the recording into the right context, additional information is collected as will be described below.

– *Post Concussion (Recovery):* Finally, once a subject has been confirmed with a concussion and is going through a rehabilitation and recovery phase, additional speech recordings obtained at specific intervals of time after when the concussion occurred can provide insights into the recovery process and duration.

Additional contextual information is essential to correctly interpreting the speech recordings, e.g., recordings from subjects that are confirmed to be concussed (via an evaluation using other, traditional concussion assessment tools) can be used to study statistical significance of various acoustic features or to train machine learning algorithms to classify subjects. It is important to note, however, that correctly classifying subjects is difficult given the subtleties of many symptoms experienced and the imperfect sensitivities of existing tests. We therefore, whenever possible, attempt to obtain scores and evaluations from more than one test.

After completing the speech protocol, the subject answers whether prescription or over-the-counter medications are being taken currently, which could have an impact on the speaking performance. These are only two of many possible confounders, therefore, future versions of our data collection application are likely to ask additional questions, e.g., with respect to sleep quality, level of exhaustion, stress levels, potential intoxication, etc. When the "recording with contact" data collection is concluded, the subject is also asked when the suspected injury occurred (since time between injury and recording may impact the evaluation results), and the location of the head impact if known (front, back, right, left, rotational).

In addition, at anytime (before, after, and independent from taking the reading test) the subject, athletic trainer, or physician can provide additional information that will be relevant for analysis, as shown in Table 4.

Figure 5 shows screenshots of some of the screens of the reading test as implemented for the Apple iPad device. The reading tests for categories 1 and 3 (words and sentences) are straightforward and the subject is simply shown what to read into the microphone in the center of the screen. At anytime, the subject can abort (and retake) a test. Test 2 requires the subject to put emphasis on different parts of the sentence. Toward that end, the subject is instructed how to read the sentence ("read the sentence, saying the word in CAPS louder"); the sentence then shows the word to emphasize in uppercase letters with an arrow and the annotation "Louder" further ensuring that the test instructions are clear (second screenshot in Figure 5). The most challenging tests

TABLE 4. Subject questionnaire.

Question	Options/Answers
Injury evaluated by:	Physician Athletic trainer
Diagnosis:	Concussed Not concussed
Classification grade:	Grade I (mild) Grade II (moderate Grade III (severe)
Other tests performed (including score):	Brain imaging ImPACT Axon King Devick SAC SCAT2/SCAT3 BESS Physicals Other (specify)
Symptoms:	headache "pressure in head" neck pain nausea or vomiting dizziness blurred vision balance problems sensitivity to light sensitivity to noise feeling slowed down feeling like "in a fog" "don't feel right" difficulty concentrating difficulty remembering fatigue or low energy confusion drowsiness trouble falling asleep more emotional irritability sadness nervous or anxious

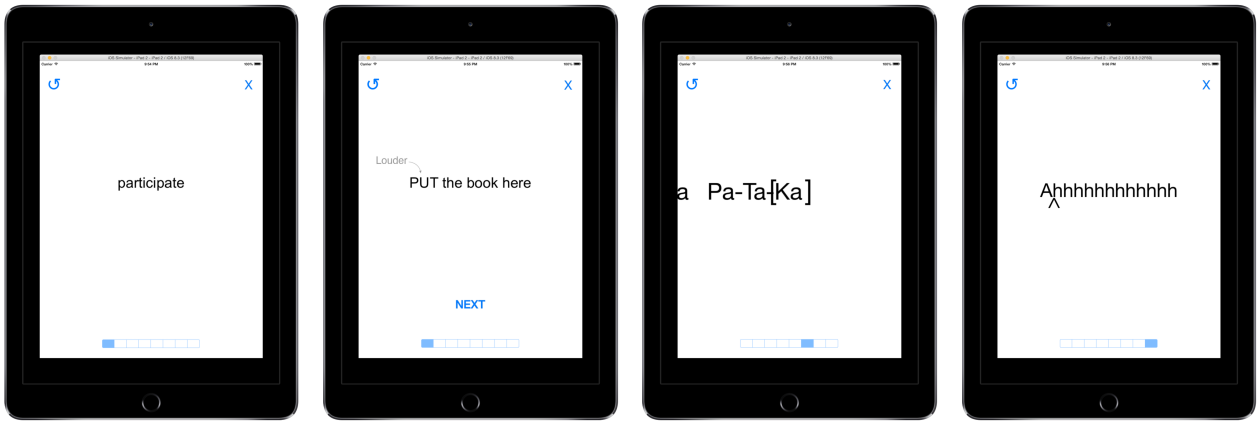

FIGURE 5. Test 1: Screenshots showing some of the reading tests.

are tests 4, 5, and 6 (sequential and alternating motion rates); for these, the subject is first instructed to slowly read the sequence shown on the screen (''pa'', ''ka'', or ''pa-ta-ka''), where the sequence scrolls from right to left, with brackets in the center of the screen indicating what to read (third screenshot in Figure 5). This is repeated at a somewhat faster speed and then a third time, where the subject is asked to take a deep breath and read as fast as possible (while the text scrolls from right to left at a high speed). Finally, test 7 requires the subject to sustain the ''aahhhh'' sound for several seconds; in this part of the reading test, the sequence is shown on the screen with an arrow underneath scrolling from left to right, indicating how long to read the sound (final screenshot in Figure 5).

As shown in Figure 4, the speech recording undergoes an ''SNR calculation and threshold comparison'' phase that is intended to ensure that the recording is of sufficient quality (as described in Section IV-B). If the quality is sufficient, the recording is then transmitted to a remote server for further processing (a diagnostic application without server/network reachability could perform these processing steps also directly on the mobile device). If the quality is not sufficient, the subject is requested to repeat the data collection, possibly adjusting the microphone, speaking louder, or moving to a quieter location. In our case study, for simplicity, we discarded an entire test category if the SNR was too low on average for that category. As a consequence, on average 1.5 tests (out of the seven categories) were discarded per subject. A less rigorous approach could discard only those portions of a test category that have a low SNR, thereby increasing the amount of data usable for analysis.

Once a recording of sufficient quality has been obtained, it is processed as shown in Figure 6. First, as described in Section IV-C, automatic speech recognition techniques are used to detect boundaries of linguistic entities. This is followed by a feature extraction step as described in Section IV-D. In our current implementation, all features are extracted using the open-source CMU Sphinx speech processing toolkit [74]–[76]. If the features are extracted for research purposes only, they can now be evaluated and analyzed. For diagnostic purposes, the features would then be compared to the features from the same subject's baseline recording (stored in a database) and the subject could be classified as either concussed or not.

B. DATA COLLECTION

We used the application described in the previous section to perform a data collection during August and December 2014, with the goal of obtaining sufficient recordings to analyze the acoustic features described in Section IV-D. We performed a population-based case-control study composed of high school and collegiate athletes participating in sports with high concussion rates. In total, 47 schools in the Midwest (Illinois, Indiana, Michigan) and Pennsylvania agreed to participate in this effort; note that neither subjects nor physicians and athletic trainers (ATs) received incentives and all participation was voluntary. Each school was visited by one of our group members to train physicians and ATs in the use of the application and to perform baseline testing at the beginning of the school year or athletic season. Overall, more than 2,500 youth athletes enrolled in our study and were baselined. During baseline testing, the subjects were asked to fill out a questionnaire, including information such as age, gender, concussion history, and other current or prior health conditions. All files associated with an athlete were stored on the device tagged by a one-way hash applied to the athlete's name for unique, but anonymized identification. Since the same mobile devices were used to collect recordings from multiple subjects, a roster management system was also added to the application to make it easier for the physicians or ATs to administer the test and to ensure that speech recordings and other data are associated with the appropriate subjects. After each athletic event (training session or competition), ATs randomly selected a few of the subjects for testing. In addition, whenever a concussion was suspected or

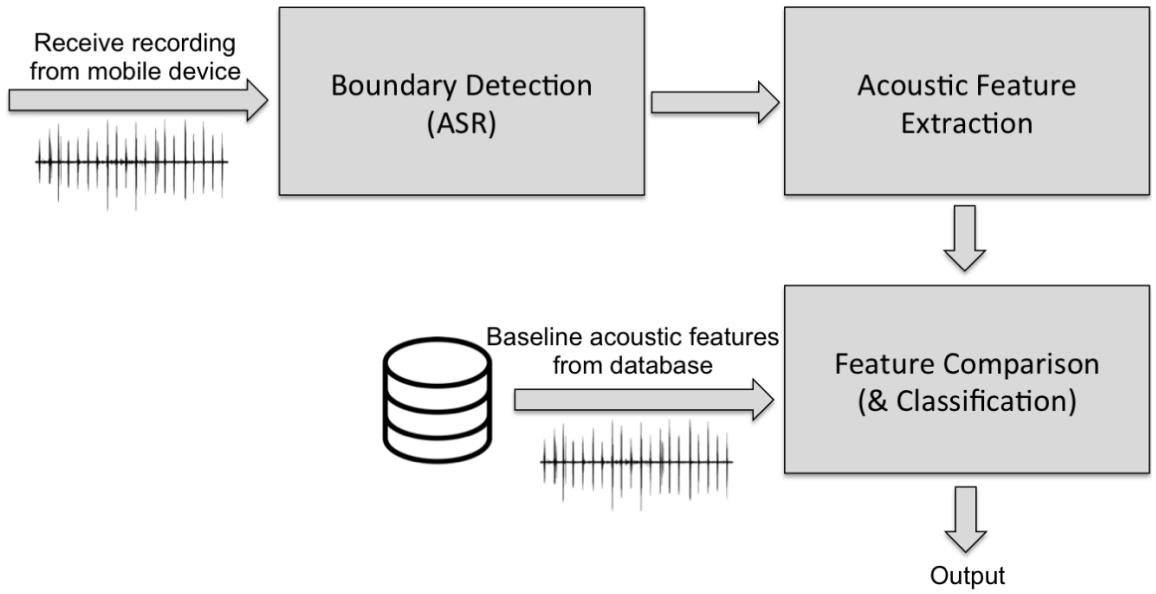

FIGURE 6. Server-side architecture of data collection application.

confirmed (e.g., by a physician or using traditional concussion testing tools), the test was also administered.

The subjects used for our study are summarized in Table 5. Note that out of the 2,500 subjects, since students were selected randomly for post-baseline data collection, many subjects have only provided a baseline recording and can therefore not be used for the analysis. From the remaining subjects, i.e., subjects with at least two recordings (baseline and an additional recording at a later time), several more were eliminate due to noise and other quality issues. In total, for the remainder of this paper, we focused on 580 subjects, 95 of them with a confirmed concussion and 485 control subjects. The male:female ratio was about 4.6:1, which was as expected given that many contact sports are ''male-centric''. Table 5 also summarizes a few other relevant information that was captured, e.g., the questionnaire asked subjects if they have any current orthodontic treatment, so that we know if the participants might be under the influence of tooth or jaw correction (dental) treatment, e.g., braces etc., as this may influence the way they speak. They were also asked if they have any speech impediments such as stuttering and learning disorders such as dyslexia.

C. DATA ANALYSIS

Given the data set obtained through the data collection study with the youth athletes, the next step is to analyze the various acoustic features for their suitability as concussion biomarkers. One of the main challenges of our data set is the risk of over-fitting, i.e., our data set has a large number of parameters relative to the number of observations. Since we are investigating 38 different vocal features, logistic regression modeling, which is a standard method of predicting and explaining a binary response variable, was used to identify

TABLE 5. Study participants.

	All	Control	Concussed
Total	580	485	95
Males	477	398	79
Females	103	87	16
Mean Age (years)	16.6	16.4	17.5
Maximum Age (years)	24	22	24
Minimum Age (years)	14	14	14
Learning Disability	11	10	1
Neurological Disorder	52	39	13
Orthodontic Treatment	45	44	1
Speech Impediment	10	8	2
Prescription Medication	87	72	15

the most promising features. First, we standardize the data set such that each feature is Gaussian with mean $\mu = 0$ and standard deviation $\sigma = 1$; then the standard (or z-scores) are computed as follows:

$$z = \frac{x - \mu}{\sigma}. \tag{3}$$

Next, we classify the data sets as either concussed or non-concussed to complete the reduction to a regression problem. The data set still has many ''holes'', because most recordings were imperfect, i.e., even though their overall SNR was satisfactory to accept a recording, various components of a test may still have to be discarded due to excessive noise. Therefore, the data set needs to be filled in using a data imputation method. Toward this end, we use the Sparco toolbox,[5] which is an open-source framework in Matlab, used for testing and benchmarking algorithms for sparse reconstruction.

After imputation, we now have a normalized data matrix and we use elastic net regularization for generalized

[5]http://www.cs.ubc.ca/labs/scl/sparco/

TABLE 6. Lasso predictor matrix.

Test Category and Acoustic Feature	Predictor Value
test1_Time_Average_Duration	**0.1367**
test1_Time_sDev_Duration	0
test2-PUT_Time_Stressed Word Duration	0
test2-PUT_Time_Stress Pause	0
test2-PUT_Pitch_F0 Movement	0
test2-PUT_Pitch_F0 Rate	0
test2-PUT_Amp_Intensity Deviation	0
test2-BOOK_Time_Stressed Word Duration	**-0.08**
test2-BOOK_Time_Stress Pause	0
test2-BOOK_Pitch_F0 Movement	0
test2-BOOK_Pitch_F0 Rate	0
test2-BOOK_Amp_Intensity Deviation	0
test2-HERE_Time_Stressed Word Duration	0
test2-HERE_Time_Stress Pause	0
test2-HERE_Pitch_F0 Movement	**-0.1439**
test2-HERE_Pitch_F0 Rate	0
test2-HERE_Amp_Intensity Deviation	**0.071**
test3_Time_SSrdur	0
test3_Time_SSpdur	0
test4_Time_Average_DDK Period	**0.3716**
test4_Time_Average_DDK Rate	0
test4_Time_sDev_DDK Period	**0.1608**
test4_Time_CV_DDK Period	0
test4_Amp_sDev_DDK Peak Intensity	0
test4_Amp_CV_DDK Peak Intensity	0
test6_Time_Average_DDK Period	0
test6_Time_Average_DDK Rate	**-0.121**
test6_Time_sDev_DDK Period	0
test6_Time_CV_DDK Period	0
test6_Amp_sDev_DDK Peak Intensity	0
test6_Amp_CV_DDK Peak Intensity	**-0.3782**
test7_Pitch_Average F0	**-0.0449**
test7_Pitch_sDev F0	0
test7_Pitch_F0 Range	0
test7_Pitch_CV vF0	0
test7_Amp_sDev F0	0
test7_Amp_CV vF0	**-0.0681**
test7_Pitch_Jitter	0

linear model regression using Lasso GLM (Generalized Linear Models) in Matlab,[6] which combines the L1 and L2 penalties of the Lasso and Ridge regression methods [77]. Lasso automatically selects the more relevant features and discards the others, whereas Ridge regression never fully discards any features. Elastic net regularization is kind of a hybrid of Ridge regression and Lasso regularization; like Lasso, elastic net regularization can generate reduced models by generating zero-valued coefficients. Elastic net regularization was primarily chosen because empirical studies suggest that it can outperform Lasso on data with highly correlated predictors [78]. For an α strictly between 0 and 1 and a nonnegative λ, elastic net regularization solves the following problem:

$$\min_{\beta_0,\beta}\left(\frac{1}{N}Deviance(\beta_0,\beta)+\lambda P_\alpha(\beta)\right), \tag{4}$$

where

$$P_\alpha(\beta)=\frac{(1-\alpha)}{2}||\beta||_2^2+\alpha||\beta||_1 \tag{5}$$

$$=\sum_{j=1}^{p}\left(\frac{(1-\alpha)}{2}\beta_j^2+\alpha|\beta_j|\right). \tag{6}$$

[6]http://www.mathworks.com/help/stats/lasso.html

When $\alpha = 1$, elastic net regularization behaves like Lasso; for other values of α, the penalty term $P_\alpha(\beta)$ interpolates between the L1 norm of β and the squared L2 norm of β. As α shrinks toward 0, elastic net regularization approaches the behavior of Ridge regression. To avoid overfitting, we cannot have fewer than 10 samples per feature given our 95 concussed recordings, i.e., in order to obtain stable results for logistic regression modeling, the data must contain at least 10 events (i.e., concussions) for every predictor variable included in the model. We are further using a 10-fold cross validation for our evaluations. In summary, using elastic net regularization (or Lasso GLM), we minimized overfitting of our data set and are now able to obtain the maximum-likelihood fitted coefficients, and thereby predictors, for the concussed/control classification problem. Table 6 shows the acoustic features (including the test categories from which they were extracted) that were evaluated and their resulting predictor values. The features shown in bold (and with non-zero predictor value) are the ten features that were selected by the elastic net regularization approach. Using these values, we can now determine an ROC curve as displayed in Figure 7, showing the true and false positives for the combination of the ten features described above. Assuming that we want to obtain a false positive rate of no more than 30%, the corresponding sensitivity of 70% can be obtained. While these results clearly show a correlation between concussions and some of the acoustic features investigated, these results are at this point insufficient to develop a speech-based concussion diagnostic tool. However, these preliminary investigations have ignored a variety of characteristics of the data set, which require further investigation, including the impact of confounders, the impact of the time interval between concussive event and recording, the various levels of severity of concussions, the number and severity of the symptoms experienced by the subject, the location of the concussive hit on the head, the results of other concussion assessment tests performed, etc. Using careful data stratification and regression modeling, we will further investigate these contextual information to maximize the confidence

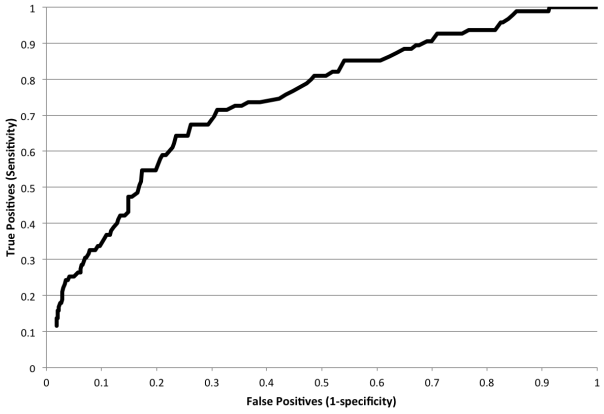

FIGURE 7. Sensitivity and specificity.

IEEE Access
The journal for rapid open access publishing

in our results and also to detect additional patterns in the speech data.

VI. DISCUSSION

The focus of the data collection described in the previous section was on obtaining a better understanding of the relationship between brain injuries and speech, which will be essential for the future development of diagnostic and assessment tools based on speech. As a diagnostic tool, a speech assessment app developed for a mobile device can be used directly at the sidelines of a sport event whenever a concussion is suspected or in the locker rooms after a sport event, e.g., to perform routine assessment. The data collections described in the previous section were performed either outdoors (at the sidelines) or indoors in a gym or locker room and therefore reflect the same environments (e.g., crowd noise) we expect to encounter when using the app for actual real-time diagnosis. A number of variables require further investigation, e.g., the optimal time between suspected concussive hit and test is yet unknown. In our data sets, concussed athletes were tested about 30 minutes (on average) after a concussive hit, with the shortest time interval being 1 minute and the largest 4 hours. While our initial results indicate that a concussion's impact on speech appears to be immediate, in order to increase accuracy, a certain minimum delay (rest) may be required to minimize the impact of exhaustion or shortness of breath.

VII. CONCLUSIONS AND FUTURE WORK

This paper describes a series of challenges experienced in the design and development of a speech-based data collection tool, as well as the use of this tool for a large-scale data collection effort among youth athletes. Unlike many other types of data, speech collection, processing, and analysis pose various unique challenges that require careful design choices and evaluations to ensure that data collection will provide recordings of the highest quality possible. While speech has received an increasing amount of attention as potential health biomarker in the recent past, further analyses remain to be done to validate its potential as biomarker and assessment tool. The focus in prior work has been on demonstrating the relationships between neurodevelopmental and neurodegenerative conditions and various vocal features. The work presented in this paper not only focuses on a larger and much more varied combination of vocal features than prior work, but also addresses the challenges in building and designing appropriate speech data collection tools and ultimately speech-based diagnostic and assessment tools. While our current results are encouraging and indicative of a strong link between brain injury and speech, there are challenges that remain to be addressed. For example, as previously mentioned, the enormous data set resulting from our 2014 data collection still requires further data cleansing and processing to ensure that analysis will provide the most reliable insights possible. Statistical analysis and machine learning research has resulted in various tools to evaluate the significance of various features and future work will continue to evaluate the most promising techniques given the nature of our data. The list of acoustic features used in this work is already 38 items long, but is still expected to grow, e.g., we are exploring techniques to produce a "similarity score" between two recordings (baseline and post-event), which may further lead to new insights. Finally, the ultimate goal will be to design and develop diagnostic tools based on our insights and to expand this research into other areas of neurology besides concussions.

ACKNOWLEDGMENT

The authors would like to thank Shane McQuillan (Contect, Inc.), Tomas Collins (Contect, Inc.), and Konrad Kording (Rehabilitation Institute of Chicago) for their support in the data collection and data analysis phases of the project.

REFERENCES

[1] J. Langlois, W. Rutland-Brown, and M. M. Wald, "The epidemiology and impact of traumatic brain injury: A brief overview," *J. Head Trauma Rehabil.*, vol. 21, no. 5, pp. 375–378, 2006.

[2] L. E. Goldstein *et al.*, "Chronic traumatic encephalopathy in blast-exposed military veterans and a blast neurotrauma mouse model," *Sci. Transl. Med.*, vol. 4, no. 134, p. 134ra60, May 2012.

[3] D. M. Sosin, D. J. Thurman, and J. E. Sniezek, "Incidence of mild and moderate brain injury in the United States, 1991," *Brain Injury*, vol. 10, no. 1, pp. 47–54, 1996.

[4] A. J. Ryan, "Intracranial injuries resulting from boxing," *Clin. Sports Med.*, vol. 17, no. 1, pp. 155–168, Jan. 1998.

[5] M. Kaste, T. Kuurne, J. Vilkki, K. Katevuo, K. Sainio, and H. Meurala, "Is chronic brain damage in boxing a hazard of the past?" *Lancet*, vol. 2, no. 8309, pp. 1186–1188, Nov. 1982.

[6] P. N. Nemetz *et al.*, "Traumatic brain injury and time to onset of Alzheimer's disease: A population-based study," *Amer. J. Epidemiol.*, vol. 149, no. 1, pp. 32–40, Jan. 1999.

[7] B. L. Plassman *et al.*, "Prevalence of dementia in the United States: The aging, demographics, and memory study," *Neuroepidemiology*, vol. 29, nos. 1–2, pp. 125–132, 2007.

[8] B. E. Gavett, R. A. Stern, and A. C. McKee, "Chronic traumatic encephalopathy: A potential late effect of sport-related concussive and subconcussive head trauma," *Clin. Sports Med.*, vol. 30, no. 1, pp. 179–188, Jan. 2011.

[9] J. S. Delaney, F. Abuzeyad, J. A. Correa, and R. Foxford, "Recognition and characteristics of concussions in the emergency department population," *J. Emergency Med.*, vol. 29, no. 2, pp. 189–197, 2005.

[10] J. Resch *et al.*, "ImPact test-retest reliability: Reliably unreliable?" *J. Athletic Training*, vol. 48, no. 4, pp. 506–511, Jul./Aug. 2013.

[11] K. U. Rani and M. S. Holi, "Analysis of speech characteristics of neurological diseases and their classification," in *Proc. 3rd Int. Conf. Comput. Commun. Netw. Technol. (ICCCNT)*, Coimbatore, India, Jul. 2012, pp. 1–6.

[12] S. Sapir, L. Ramig, and C. Fox, "Assessment and treatment of the speech disorder in Parkinson disease," *Commun. Swallowing Parkinson's Disease*, pp. 89–122, 2011.

[13] O. Geman, "Data processing for Parkinson's disease: Tremor, speech and gait signal analysis," in *Proc. IEEE Int. E-Health Bioeng. Conf.*, Iasi, Romania, Nov. 2011, pp. 1–4.

[14] A. M. Goberman, M. Blomgren, and E. Metzger, "Characteristics of speech disfluency in Parkinson disease," *J. Neurolinguistics*, vol. 23, no. 5, pp. 470–478, 2010.

[15] P. Zwirner, T. Murry, and G. E. Woodson, "Phonatory function of neurologically impaired patients," *J. Commun. Disorders*, vol. 24, no. 4, pp. 287–300, 1991.

[16] W. Ziegler and D. von Cramon, "Spastic dysarthria after acquired brain injury: An acoustic study," *Int. J. Lang. Commun. Disorders*, vol. 21, no. 2, pp. 173–187, 1986.

[17] D. G. Theodoros, B. E. Murdoch, and H. J. Chenery, "Perceptual speech characteristics of dysarthric speakers following severe closed head injury," *Brain Injury*, vol. 8, no. 2, pp. 101–124, Feb./Mar. 1994.

[18] N. K. Madigan, J. DeLuca, B. J. Diamond, G. Tramontano, and A. Averill, "Speed of information processing in traumatic brain injury: Modality-specific factors," *J. Head Trauma Rehabil.*, vol. 15, no. 3, pp. 943–956, 2000.

[19] A. D. Hinton-Bayre, G. Geffen, and K. McFarland, "Mild head injury and speed of information processing: A prospective study of professional rugby league players," *J. Clin. Experim. Neuropsychol.*, vol. 19, no. 2, pp. 275–289, 1997.

[20] M. H. Davis, I. S. Johnsrude, A. Hervais-Adelman, K. Taylor, and C. McGettigan, "Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences," *J. Experim. Psychol.*, vol. 134, no. 2, pp. 222–241, 2005.

[21] M. O. Krause, "The effects of brain injury and talker characteristics on speech processing in a single-talker interference task," Ph.D. dissertation, Univ. Minnesota, Minneapolis, MN, USA, Jul. 2011.

[22] M. Falcone, N. Yadav, C. Poellabauer, and P. Flynn, "Using isolated vowel sounds for classification of mild traumatic brain injury," in *Proc. 38th IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Vancouver, BC, Canada, May 2013, pp. 7577–7581.

[23] D. J. Gelb, E. Oliver, and S. Gilman, "Diagnostic criteria for Parkinson disease," *Arch. Neurol.*, vol. 56, no. 1, pp. 33–39, Jan. 1999.

[24] I. Ungurean and N.-C. Gaitan, "Speech analysis for medical predictions based on cell broadband engine," in *Proc. 20th Eur. Signal Process. Conf. (EUSIPCO)*, Bucharest, Romania, Aug. 2012, pp. 1733–1736.

[25] T. Bocklet, E. Noth, G. Stemmer, H. Ruzickova, and J. Rusz, "Detection of persons with Parkinson's disease by acoustic, vocal, and prosodic analysis," in *Proc. IEEE Workshop Autom. Speech Recognit. Understand. (ASRU)*, Dec. 2011, pp. 478–483.

[26] A. Maier *et al.*, "PEAKS—A system for the automatic evaluation of voice and speech disorders," *Speech Commun.*, vol. 51, no. 5, pp. 425–437, 2009.

[27] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 4, pp. 1015–1022, Apr. 2009.

[28] K. Chenausky, J. MacAuslan, and R. Goldhor, "Acoustic analysis of PD speech," *Parkinson's Disease*, vol. 2011, Jun. 2011, Art. ID 435232.

[29] A. K. Ho, R. Iansek, C. Marigliani, J. L. Bradshaw, and S. Gates, "Speech impairment in a large sample of patients with Parkinson's disease," *Behavioural Neurol.*, vol. 11, no. 3, pp. 131–137, 1998.

[30] M. A. Little, P. E. McSharry, S. J. Roberts, D. A. E. Costello, and I. M. Moroz, "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection," *Biomed. Eng. OnLine*, vol. 6, p. 23, Jun. 2007.

[31] J. R. Duffy, *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management*, 3rd ed. St. Louis, MO, USA: Mosby, 2005.

[32] F. L. Darley, A. E. Aronson, and J. R. Brown, *Motor Speech Disorders*. Philadelphia, PA, USA: Saunders, 1975.

[33] C. Mackenzie, "Dysarthria in stroke: A narrative review of its description and the outcome of intervention," *Int. J. Speech-Lang. Pathol.*, vol. 13, no. 2, pp. 125–136, 2011.

[34] B. M. Landry, E. J. Choe, S. McCutcheon, and J. A. Kientz, "Post traumatic stress disorder: Issues and opportunities," in *Proc. 1st ACM Int. Health Informat. Symp.*, Arlington, VA, USA, Nov. 2010, pp. 780–789.

[35] American Psychiatric Association, *Diagnostic and statistical manual of mental disorders: DSM-IV*, 4th ed. Washington, DC, USA: APA, 1994.

[36] M. Sharda *et al.*, "Sounds of melody—Pitch patterns of speech in autism," *Neurosci. Lett.*, vol. 478, no. 1, pp. 42–45, 2010.

[37] M. Valicenti-McDermott, K. Hottinger, R. Seijo, and L. Shulman, "Age at diagnosis of autism spectrum disorders," *J. Pediatrics*, vol. 161, no. 3, pp. 554–556, 2012.

[38] G. Dawson, S. Rogers, J. Munson, M. Smith, J. Winter, and J. Greenson, "Randomized, controlled trial of an intervention for toddlers with autism: The Early Start Denver Model," *Pediatrics*, vol. 125, no. 1, pp. e17–e23, 2010.

[39] A. Fort and C. Manfredi, "Acoustic analysis of newborn infant cry signals," *Med. Eng. Phys.*, vol. 20, no. 6, pp. 432–442, 1998.

[40] C. Manfredi, L. Bocchi, S. Orlandi, L. Spaccaterra, and G. P. Donzelli, "High-resolution cry analysis in preterm newborn infants," *Med. Eng. Phys.*, vol. 31, no. 5, pp. 528–532, 2009.

[41] S. J. Sheinkopf, J. M. Iverson, M. L. Rinaldi, and B. M. Lester, "Atypical cry acoustics in 6-month-old infants at risk for autism spectrum disorder," *Autism Res.*, vol. 5, no. 5, pp. 331–339, Oct. 2012.

[42] B. Mampe, A. D. Friederici, A. Christophe, and K. Wermke, "Newborns' cry melody is shaped by their native language," *Current Biol.*, vol. 19, no. 23, pp. 1994–1997, Dec. 2009.

[43] K. Wermke, W. Mende, C. Manfredi, and P. Bruscaglioni, "Developmental aspects of infant's cry melody and formants," *Med. Eng. Phys.*, vol. 24, nos. 7–8, pp. 501–514, 2002.

[44] K. Lind and K. Wermke, "Development of the vocal fundamental frequency of spontaneous cries during the first 3 months," *Int. J. Pediatric Otorhinolaryngol.*, vol. 64, no. 2, pp. 97–104, 2002.

[45] J. Brisson, K. Martel, J. Serres, S. Sirois, and J.-L. Adrien, "Acoustic analysis of oral productions of infants later diagnosed with autism and their mother," *Infant Mental Health J.*, vol. 35, no. 3, pp. 285–295, 2014.

[46] G. Kiss, J. P. H. van Santen, E. T. Prud'hommeaux, and L. M. Black, "Quantitative analysis of pitch in speech of children with neurodevelopmental disorders," in *Proc. 13th Annu. Conf. Int. Speech Commun. Assoc. (INTERSPEECH)*, Portland, OR, USA, Sep. 2012, pp. 1343–1346.

[47] S. Peppe, J. Cleland, F. Gibbon, A. O'Hare, and P. M. Castilla, "Expressive prosody in children with autism spectrum conditions," *J. Neurolinguistics*, vol. 24, no. 1, pp. 41–53, 2011.

[48] M. Brenner and J. R. Cash, "Speech analysis as an index of alcohol intoxication—The Exxon Valdez accident," *Aviation, Space, Environ. Med.*, vol. 62, no. 9, pp. 893–898, Sep. 1991.

[49] F. Schiel and C. Heinrich, "Laying the foundation for in-car alcohol detection by speech," in *Proc. INTERSPEECH*, Brighton, U.K., 2009, pp. 983–986.

[50] D. Bone, M. P. Black, M. Li, A. Metallinou, S. Lee, and S. S. Narayanan, "Intoxicated speech detection by fusion of speaker normalized hierarchical features and GMM supervectors," in *Proc. 12th Annu. Conf. Int. Speech Commun. Assoc. (INTERSPEECH)*, Aug. 2011, pp. 3217–3220.

[51] M. Levit, R. Huber, A. Batliner, and E. Noeth, "Use of prosodic speech characteristics for automated detection of alcohol intoxication," in *Proc. Workshop Prosody Speech Recognit. Understand.*, Red Bank, NJ, USA, Oct. 2001, pp. 22–24.

[52] *The Global Burden of Disease: 2004 Update*, World Health Org., Geneva, Switzerland, 2004.

[53] R. Layard, "The case for psychological treatment centres," *Brit. Med. J.*, vol. 332, no. 7548, pp. 1030–1032, 2006.

[54] K. Kasai *et al.*, "Impaired cortical network for preattentive detection of change in speech sounds in schizophrenia: A high-resolution event-related potential study," *Amer. J. Psychiatry*, vol. 159, no. 4, pp. 546–553, Apr. 2002.

[55] V. Rapcan, S. D'Arcy, S. Yeap, N. Afzal, J. Thakore, and R. B. Reilly, "Acoustic and temporal analysis of speech: A potential biomarker for schizophrenia," *Med. Eng. Phys.*, vol. 32, no. 9, pp. 1074–1079, Nov. 2010.

[56] L. A. Ramig, R. C. Scherer, I. R. Titze, and S. P. Ringel, "Acoustic analysis of voices of patients with neurologic disease: Rationale and preliminary data," *Ann. Otol., Rhinol., Laryngol.*, vol. 97, no. 2, pp. 164–172, Mar./Apr. 1988.

[57] H. Halpern and R. Goldfarb, *Language and Motor Speech Disorders in Adults*, 3rd ed. Boston, MA, USA: Jones & Bartlett, 2012.

[58] M. Okada, "Measurement of speech patterns in neurological disease," *Med. Biol. Eng. Comput.*, vol. 21, no. 2, pp. 145–148, 1983.

[59] T. M. Sullivan and R. M. Stern, "Multi-microphone correlation-based processing for robust speech recognition," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 1993, pp. 91–94.

[60] H. Hermansky, N. Morgan, and H.-G. Hirsch, "Recognition of speech in additive and convolutional noise based on RASTA spectral processing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 1993, pp. 83–86.

[61] N. Yadav, L. Daudet, C. Poellabauer, and P. Flynn, "Noise management in mobile speech based health tools," in *Proc. IEEE EMBS Special Topic Conf. Healthcare Innov. Point-Care Technol.*, Seattle, WA, USA, Oct. 2014, pp. 335–338.

[62] R. Hook and T. A. Jeeves, "'Direct search' solution of numerical and statistical problems," *J. ACM*, vol. 8, no. 2, pp. 212–229, 1961.

[63] X. Huang, A. Acero, and H. W. Hon, *Spoken Language Processing*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2001.

[64] L. Deng and D. O'Shaughnessy, *Speech Processing—A Dynamic and Optimization-Oriented Approach*. New York, NY, USA: Marcel Dekker, Jun. 2003.

[65] D. Huggins-Daines, M. Kumar, A. Chan, A. W. Black, M. Ravishankar, and A. I. Rudnicky, "Pocketsphinx: A free, real-time continuous speech recognition system for hand-held devices," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, vol. 1. Toulouse, France, May 2006, pp. 185–188.
[66] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.
[67] D. B. Paul and J. M. Baker, "The design for the Wall Street Journal-based CSR corpus," in *Proc. Workshop Speech Natural Lang.*, Harriman, NY, USA, Feb. 1992, pp. 357–362.
[68] D. Graff, J. Garofolo, J. Fiscus, W. Fisher, and D. Pallett. (1996). 1996 English broadcast news speech (HUB4). Lingustic Data Consortium, Philadelphia, PA, USA. [Online]. Available: https://catalog.ldc.upenn.edu/LDC97S44
[69] J.-L. Gauvain and C.-H. Lee, "Maximum *a posteriori* estimation for multivariate Gaussian mixture observations of Markov chains," *IEEE Trans. Speech Audio Process.*, vol. 2, no. 2, pp. 291–298, Apr. 1994.
[70] C. J. Leggetter and P. C. Woodland, "Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models," *Comput. Speech Lang.*, vol. 9, no. 2, pp. 171–185, Apr. 1995.
[71] G. E. Peterson and H. L. Barney, "Control methods used in a study of the vowels," *J. Acoust. Soc. Amer.*, vol. 24, no. 2, pp. 175–184, 1952.
[72] J. Hillenbrand, L. A. Getty, M. J. Clark, and K. Wheeler, "Acoustic characteristics of American English vowels," *J. Acoust. Soc. Amer.*, vol. 97, no. 5, pp. 3099–3111, May 1995.
[73] M. Vasilakis and Y. Stylianou, "Voice pathology detection based eon short-term jitter estimations in running speech," *Folia Phoniatrica Logopaedica*, vol. 61, no. 3, pp. 153–170, 2009.
[74] D. D. Deliyski, H. S. Shaw, and M. K. Evans, "Influence of sampling rate on accuracy and reliability of acoustic voice analysis," *Logopedics Phoniatrics Vocol.*, vol. 30, no. 2, pp. 55–62, 2005.
D. Huggins-Daines, M. Kumar, A. Chan, A. W. Black, M. Ravishankar, and A. I. Rudnicky, "Pocketsphinx: A free, real-time continuous speech recognition system for hand-held devices," in *Proc. Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2006, pp. 185–188.
[75] K.-F. Lee, H.-W. Hon, and R. Reddy, "An overview of the SPHINX speech recognition system," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 1, pp. 35–45, Jan. 1990.
[76] A. Varela, H. Cuayáhuitl, and J. A. Nolazco-Flores, "Creating a Mexican Spanish version of the CMU Sphinx-III speech recognition system," in *Proc. 8th Iberoamer. Congr. Pattern Recognit. (CIARP)*, 2003, pp. 251–258.
[77] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970.
[78] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. Roy. Statist. Soc., B*, vol. 67, no. 2, pp. 301–320, 2005.

CHRISTIAN POELLABAUER (S'97–M'04–SM'09) received the Diplom Ingenieur degree in computer science from the Vienna University of Technology, Austria, in 1998, and the Ph.D. degree in computer science from the Georgia Institute of Technology, Atlanta, GA, in 2004. He is currently an Associate Professor with the Department of Computer Science and Engineering, University of Notre Dame. He has authored over 100 papers in his research areas, and co-authored a textbook in wireless sensor networks. His research interests are in the areas of wireless sensor networks, mobile computing, ad-hoc networks, pervasive computing, and mobile healthcare systems. His research has received funding through the National Science Foundation (including the CAREER Award in 2006), the National Institutes of Health, the Department of Education, the Moore Family Foundation, the Army Research Office, the Office of Naval Research, IBM, Intel, Toyota, GE Health, the National Football League, Ford Research, Serim Research Corporation, and Motorola Labs.

NIKHIL YADAV received the B.Eng. degree from the National University of Lesotho, in 2004, and the M.S. degree in computer engineering from the University of Florida, in 2007. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering, University of Notre Dame. He gained work experience with iThemba Labs, Tech Source, Inc., and Cloverapps, LLC. His research interests are in the development of healthcare technologies, in particular, focusing on speech analysis for neurological conditions and brain injuries.

LOUIS DAUDET received the B.A. degree in computer science, mathematics, and physics from Augustana College, IL, in 2006. He is currently pursuing the Ph.D. degree in computer science with the University of Notre Dame under his adviser of Dr. Poellabauer. After three years, he worked in the field of computer science in France, including two years with the Sopra Group, where he was involved in the development of their payroll solution. His research is in the development of healthcare systems on mobile devices, which includes work on voice analysis, speech recognition, speech data collection, and data mining of vocal features.

SANDRA L. SCHNEIDER received the B.S. degree in speech-language pathology from Western Michigan University, the M.S. degree in speech and hearing sciences from Vanderbilt University, and the Ph.D. degree in communication sciences and disorders from Northwestern University. She held a post-doctoral fellowship in medical speech pathology with the Mayo Graduate School of Medicine. After she held faculty positions with Ohio State University and Vanderbilt University, she is currently a Professor of Communicative Sciences and Disorders with Saint Mary's College, and an Adjunct Research Professor with the Department of Computer Science and Engineering, University of Notre Dame. Her research interests are in the area of neurogenic communication disorders (e.g., stroke, TBI, dementia, motor speech, and neurodegenerative disease processes, such as Parkinson's and ALS), which has led to multiple papers and presentations both nationally and internationally. Her current research focus is on speech as a biomarker for the detection, assessment, and treatment of neurological disorders. This work has received funding through the National Science Foundation, the NFL/GE Head Health Foundation, and the National Institutes of Health.

CARLOS BUSSO (S'02–M'09–SM'13) received the B.S. and M.S. (Hons.) degrees from the University of Chile, Santiago, Chile, in 2000 and 2003, respectively, and the Ph.D. degree from the University of Southern California, in 2008, all in electrical engineering. He is currently an Associate Professor with the Electrical Engineering Department, The University of Texas at Dallas, where he leads the Multimodal Signal Processing Laboratory. His research interests are in digital signal processing, speech and video processing, and multimodal interfaces. His current research includes the broad areas of affective computing, multimodal human–machine interfaces, modeling and synthesis of verbal and nonverbal behaviors, sensing human interaction, in-vehicle active safety system, and machine learning methods for multimodal processing. He received the ICMI Ten-Year Technical Impact Award in 2014, and the Hewlett Packard Best Paper Award at the IEEE ICME in 2011 (with J. Jain). He is the co-author of the winning paper of the Classifier Sub-Challenge Event at the Interspeech Emotion Challenge in 2009.

PATRICK J. FLYNN (S'84–M'90–SM'96–F'12) received the Ph.D. degree in computer science from Michigan State University, in 1990. He has held faculty positions with Washington State University and Ohio State University. He is currently the Duda Family Professor of Engineering with the University of Notre Dame. His research interests include computer vision, biometrics, signal and image processing, and mobile computing. He is a fellow of the International Association of Rehabilitation Professionals, and an ACM Distinguished Scientist. He was the Associate Editor-in-Chief of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, and an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, *Pattern Recognition*, and *Pattern Recognition Letters*.