Detection of Simulated Vocal Dysfunctions Using Complex sEMG Patterns

Nicholas R. Smith, Luis A. Rivera, *Student Member, IEEE*, Maria Dietrich, Chi-Ren Shyu, *Senior Member, IEEE*, Matthew P. Page, and Guilherme N. DeSouza, *Senior Member, IEEE*

***Abstract*—Symptoms of voice disorder may range from slight hoarseness to complete loss of voice; from modest vocal effort to uncomfortable neck pain. But even minor symptoms may still impact personal and especially professional lives. While early detection and diagnosis can ameliorate that effect, to date, we are still largely missing reliable and valid data to help us better screen for voice disorders. In our previous study, we started to address this gap in research by introducing an ambulatory voice monitoring system using surface electromyography (sEMG) and a robust algorithm (HiGUSSS) for pattern recognition of vocal gestures. Here, we expand on that work by further analyzing a larger set of simulated vocal dysfunctions. Our goal is to demonstrate that such a system has the potential to recognize and detect real vocal dysfunctions from multiple individuals with high accuracy under both intra and intersubject conditions. The proposed system relies on four sEMG channels to simultaneously process various patterns of sEMG activation in the search for maladaptive laryngeal activity that may lead to voice disorders. In the results presented here, our pattern recognition algorithm detected from two to ten different classes of sEMG patterns of muscle activation with an accuracy as high as 99%, depending on the subject and the testing conditions.**

***Index Terms*—Biomedical monitoring, Biomedical telemetry, Electromyography, Medical diagnosis, Source Separation, Pattern recognition.**

I. Introduction

OCCUPATIONAL voice users are at the highest risk for voice disorders largely due to the extraordinary vocal load placed on the laryngeal system while exercising their occupation [1]. Classic symptoms are hoarseness, vocal effort, and vocal fatigue, which are related to vocal hyperfunction [1]. In turn, vocal hyperfunction may lead to phonotrauma—e.g., vocal nodules—or to primary Muscle Tension Dysphonia (MTD)—i.e., excessive or dysregulated laryngeal muscular activity underlying the vocal changes [2]. So, the ability to differentiate normal and abnormal patterns of laryngeal muscular activity in daily life could improve our ability to detect and understand the pathophysiological processes leading to MTD, and thereby improving the diagnosis of this voice disorder.

In parallel to that, it is well known that as muscles contract, they undergo changes in electrical potentials, which can be monitored by electromyographic (EMG) devices. When studying a single muscle, the optimal signal to noise ratio is typically obtained when the electrodes are placed inside the muscle—a technique available in the healthcare or laboratory setting, but with limited use in people's everyday lives. A less invasive strategy is to place surface EMG (sEMG) electrodes on the skin near the muscle(s) of interest. Recently, the interest in many areas such as human–computer interfacing [3], prosthesis [4], and even voice pathology [5], [6] has fomented development of devices that can monitor muscle activity, and systems like the Delsys' Trigno, Great Lakes Neurotechnologies' BioRadio, and the Shimmer's Shimmer3 wearable sensors are just a few of the many examples in the market today.

Given the benefits of detecting MTD and the proliferation of sEMG devices in the context of other muscle-related dysfunctions (ALS, Cerebral Palsy, etc. [7]), it stands to reason that sEMG can be a powerful, noninvasive and well-suited tool also in the study of MTD. In that sense, even though reliable detection of sEMG signals in extralaryngeal muscular activity that can be associated to high risk of voice disorders is a nontrivial and scarcely investigated area, it should also be the foundation to study differences between normal and maladaptive muscular activity during voice production for speech.

In the past decade, research intensified the development of devices suitable for ambulatory monitoring of daily voice use. Commercial systems were made available to monitor vocal duration, voice intensity, and voice fundamental frequency (f_0) using accelerometers, microphones, and frequency transforms [8]–[10]. Some of these noninvasive monitoring system even provide biofeedback capabilities and have been ported to smartphone platforms [11]. However, the richness and complexity of vocal patterns during speech go well beyond what can be captured by microphones and inertial sensors. In that sense, and with limited use to extralaryngeal activity, one of the latest developments was a smartphone-based vocal health monitor in which collection of frequency and inertial data was calibrated to aerodynamic parameters, in particular glottal air flow [11]–[14].

Manuscript received February 17, 2015; revised July 3, 2015 and August 30, 2015; accepted September 24, 2015. Date of publication October 12, 2015; date of current version May 9, 2016.

N. R. Smith and L. A. Rivera are with the Vision-Guided and Intelligent Robotics Lab, University of Missouri, Columbia, MO 65211 USA, and also with the Department of Electrical and Computer Engineering, University of Missouri, Columbia, MO 65211 USA (email: nrs522@mail.missouri.edu; larmbc@mail.missouri.edu).

C. R. Shyu is with the Department of Electrical and Computer Engineering, University of Missouri, Columbia, MO 65211 USA, and also with the MU Informatics Institute, University of Missouri, Columbia, MO 65211 USA (email: shyuc@missouri.edu).

G. N. DeSouza is with the Vision-Guided and Intelligent Robotics Lab, University of Missouri, Columbia, MO 65211 USA, the Department of Electrical and Computer Engineering, University of Missouri, Columbia, MO 65211 USA, and also with the MU Informatics Institute, University of Missouri, Columbia, MO 65211 USA (email: desouzag@missouri.edu).

M. Dietrich is with the Department of Communication Science and Disorders, University of Missouri, Columbia, MO 65211 USA (email: dietrichm@health.missouri.edu).

M. P. Page is with the Otolaryngology—Head and Neck Surgery, University of Missouri, Columbia, MO 65211 USA (email: pagemp@health.missouri.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JBHI.2015.2490087

On the other hand, vocal effort is thought to be partially the result of compensatory extralaryngeal activity to produce stronger or more consistent voice during vocal fatigue [15]. Hence, even though the detection of vocal effort and fatigue as early signs of MTD is often elusive in the isolated screening or clinical setting [12], classifying sEMG signals can help to answer unique research questions about magnitude and pattern of that same extralaryngeal activity and their correlation with MTD.

Finally, while we agree with recent statements that sEMG has not reached its full potential for application to clinical and basic research in voice, speech, and swallowing [6], ambulatory monitoring of voice using sEMG on extralaryngeal muscles can be an innovative approach to advance our understanding of vocal hyperfunction and ultimately to monitor its occurrence in heavy voice users. Indeed, the broader and more pertinent issue at hand is to further research on how to best process and analyze data from any voice ambulatory monitoring system and to determine the data's clinical utility [10], [12], which is also the focus of this research, in the context of sEMG signals.

As alluded to earlier, sEMG is a less invasive strategy than its nonsurface counterpart, with more real-world practicality. However, it comes at the expense of noisier signals and exacerbated occurrences of crosstalk between adjacent electrodes. That is because the biologic functions that are subserved by muscular activity do not result from the action of a single muscle, but from the activity of several muscles working in a coordinated fashion. Moreover, when it comes to recognizing the crosstalk patterns of muscle activity in a reliable, accurate, and robust manner, much remains to be done. In our previous study [16], an ambulatory sEMG device named EMG multichannel hardware for otolaryngology (ECHO) was proposed to log sEMG data from multiple differential sEMG sensor channels. The system was connected to the anterior neck of the subject since many complex physiological motor functions underlying voice, speech, and swallowing occur within the neck. Also, the muscles in this area are located relatively close to the skin and are quite appropriate for sEMG. Our goal in [16] was to demonstrate that: 1) an ambulatory sEMG device can help to build our understanding of complex laryngeal patterns underlying voice for speech and nonspeech vocal behaviors through sEMG signals; and 2) the neck offers an excellent location to capture such signals. The new hierarchical algorithm called HiGuided-Underdetermined Source Signal Separation (HiGUSSS) was then tested for a single test subject and it achieved a classification accuracy of over 90% for six *gestures*.

So, in this paper, two new research questions have been raised: 1) whether sEMG devices can reliably associate a larger number of patterns of extralaryngeal muscle activity with voice tasks underlying speech and nonspeech behaviors; and 2) whether they can differentiate between multiple vowel sounds produced in a normal manner compared with a pressed (low air flow) manner for intra and intersubject testing. It should go without saying that the answer to these questions can lead to a method applicable clinically to the *detection* of normal and maladaptive extralaryngeal patterns associated with voice problems.

In order to answer these questions, we drastically expanded on the testing and validation of the system proposed in [16] by: 1) more than doubling the number of test subjects (ten) with different ages and genders; 2) adding new groups of different gestures with both similar and distinct patterns representing normal and simulated dysfunctional conditions; 3) testing a larger number (ten) of vocal gestures; and 4) creating different test scenarios involving intra and intersubject cases. The results presented at the end of this paper show that despite the complexity of the muscle groups on the neck, meaningful detection of vocal dysfunctions through the recognition of sEMG signals is possible, at high levels of accuracy.

II. Background and Related Work

A. Voice Disorders

As mentioned earlier, classic symptoms of behavioral voice disorders are related to vocal hyperfunction. In some individuals, vocal hyperfunction leads to phonotrauma, such as vocal nodules, while in others it leads to primary MTD [2], [17]. Excessive or disorganized extralaryngeal muscle activity and a chronic high laryngeal position during speech production are characteristic of MTD [2], [18]. Also, sEMG is a noninvasive tool that has been used in voice research. However, past research using sEMG to study hyperfunctional voice disorders was methodologically difficult to compare because of differences in inclusion criteria, sensor locations, experimental paradigms, and analysis methods [6], [19], [44]. As a result, a continued lack of systematic studies on extralaryngeal muscle activity in voice disorder (i.e., MTD) is still noted ([2], [19]) also because investigations using sEMG to study vocal hyperfunction either: 1) used clinical groups that mixed patients with and without vocal fold lesions, for example, to study relationships with neck tension palpation ratings [5], or 2) focused on patients with vocal nodules exclusively, finding that intermuscular beta coherence may be a promising indicator of vocal hyperfunction [20]. In other words, we need studies on sEMG activity that focus on understanding the excessive or disorganized extralaryngeal activity leading to symptoms of vocal effort and strain in individuals with primary MTD separately from individuals with secondary MTD—i.e., with vocal nodules, and for which MTD is a compensatory response to phonotrauma. Information on this differentiation is necessary at this stage of inquiry because different intra and extralaryngeal mechanisms were proposed for patients with primary MTD as opposed to vocal nodules (intralaryngeal, hypoadducted hyperfunction versus hyperadducted hyperfunction) [2], [17]. Also, regarding extralaryngeal activity, individuals scoring higher on introversion showed greater infrahyoid activity than submental activity during baseline speech and stressful public speaking as well as greater perceived vocal effort compared with peers with extroversion, which is in agreement with the trait theory of voice disorder's prediction that individuals with introversion may be more prone to primary MTD than vocal nodules [19], [21], [22]. The new frontier will be to test analysis methods for the data stream from extralaryngeal muscle activity to learn about normal and altered patterns during voice and nonspeech laryngeal behaviors that are linked with risk for voice disorders. Exploiting methods for ambulatory monitoring of vocal function using sEMG has great potential to help with the early detection of problems that can be elusive at early stages among professional voice users—e.g., student teachers.

As pointed out earlier, sEMG of the anterior neck is well suited to capture general information on the muscular activity of the larynx, which can be recognized as signal patterns. Presumably, excessive muscle activity or lack of variability in laryngeal movements may be related to symptoms of vocal effort and vocal fatigue, which again particularly plagues occupational voice users such as teachers [1].

In this context, measures of phonatory aerodynamic function such as air flow (L/s) and subglottal pressure (cm H_2O) during voicing give a reliable account of laryngeal valving activity during voice production [23]–[25]. A derivative measure, laryngeal airway resistance (subglottic air pressure in cm H_2O divided by air flow in L/s) is clinically relevant and used to discriminate normal and pathologic vocal function, to assess severity, and to aid management planning [25], [26]. The implications of elevated laryngeal airway resistance (e.g., decreased air flow and/or increased air pressure) would be an increased risk for vocal effort and fatigue: the most frequent vocal symptoms reported by teachers [1], [27]. Perceptually, a strained or pressed voice quality is often observed in MTD [2]. A pressed voice, characterized by a decrease in air flow, can be simulated effectively [24], and thus, can serve as an initial model for differentiating extralaryngeal sEMG patterns associated with normal versus pressed voice productions.

These studies support the pursuit of our research questions, which are again whether sEMG devices can reliably: 1) associate patterns of extralaryngeal muscle activity with voice tasks underlying speech and nonspeech behaviors (e.g., voiced sounds, throat clear); and 2) differentiate between vowel sounds produced in a normal manner compared with a pressed (low air flow) manner for intra and intersubject testing.

B. Pattern Recognition of sEMG Signals

The work in [28] introduced the idea of Guided Underdetermined Source Signal Separation (GUSSS) and the GUSSS ratio. In [28], the focus was on discriminating different muscle unit activation potential trains, or MUAPT patterns, that emerge when different gestures are performed. As many systems do, it was assumed that an sEMG sensor captures a combination of statistically independent MUAPTs due to crosstalk [29], [30]. Unlike most methods in the literature, the system in [28] relied on a single sensor. This was possible because the main characteristic of the GUSSS ratio is its ability to indicate the presence or absence of a particular signature or MUAPT pattern within a sensed sEMG signal. The term “Guided” in GUSSS refers to the fact that the sought-out signature—i.e., a previously learned signal—is “injected” into the observed signal in order to obtain a corresponding ratio. A low ratio indicates that the signature is most likely present within the sensed signal. A high ratio, on the other hand, indicates that the signature is not being detected in the signal.

Later, a framework for controlling a power wheelchair using the GUSSS method was developed and tested in [31]. The framework proposed a control system based on the recognition of hand gestures. The use of hand gestures was simply to illustrate the fact that any muscle activation pattern or signature derived from natural and repetitive muscle movements can be employed by the system. In the case of a person with severe impairment, any other muscle movement could be used instead (e.g., eyebrow movement). Compared to other systems found in the literature, which use multiple sEMG sources for classification, the method in [31] compared quite reasonably, reaching up to 92% accuracy for three gestures.

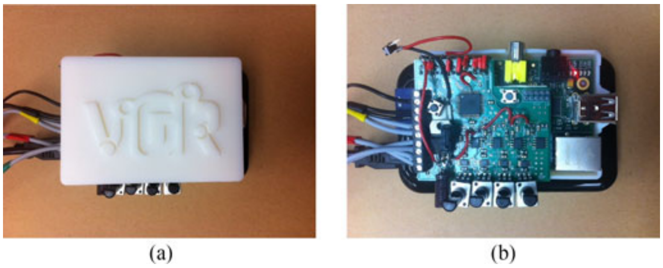


Fig. 1. ECHO device depicted from two top angles: a) with and b) without the case cover.

More recently in [32], a hierarchical system based on the GUSSS (HiGUSSS) was developed to achieve higher classification accuracy for a greater number of gestures. The HiGUSSS framework repeats the GUSSS process in parallel for tuples of prelearned signatures (e.g., doubles, triples, etc.). The reason for the use of tuples is twofold: search for multiple signatures in parallel, hence faster; and to separate similar signatures in order to avoid confusion between similar gestures, and hence, increase the success of classification even as the number of gestures increases—up to 86% accuracy for nine gestures.

As mentioned earlier, in the study presented in [16], an improved version of the HiGUSSS algorithm was applied to the detection of vocal gestures using four channels of sEMG signals collected at the anterior neck of a single subject. Six gestures (*/u/, /i/, /t/, /s/, cough,*and*throat clear*) were tested and the system achieved a classification accuracy of 85%. A small prototype device for real-time monitoring and collection of signals was also introduced. Here, we explain the design of that device and its potential application to the detection and diagnosis of voice dysfunctions.

C. Device Description

The proposed ECHO in [16] is an Otolaryngology REcording, Analysis, and Diagnostic device (OREAD) to log sEMG data from multiple differential sEMG sensor channels. One key feature of the ECHO-OREAD device is that it maintains a small form factor (8.5 cm × 6 cm × 4.5 cm) in order to be portable so that it can be used in a variety of applications. The device is connected to a rechargeable lithium-ion battery to maintain portability. Fig. 1 shows two pictures of the ECHO-OREAD device sitting on top of the rechargeable battery and Fig. 2 shows two sets of signals from all four channels captured with ECHO-OREAD for the Cough and /t/ gestures, with electrode placement as described in Section IV-A1. Next, we provide more details on the design of the ECHO-OREAD device.

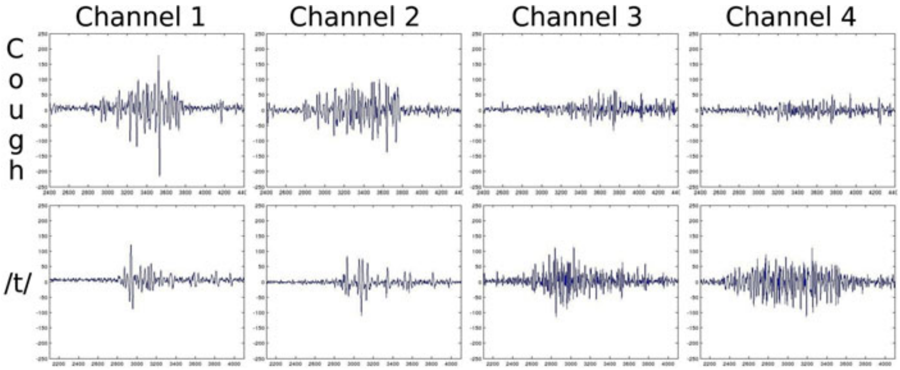


Fig. 2. Four channels of sEMG signals for gestures Cough and /t/ collected by the ECHO-OREAD.

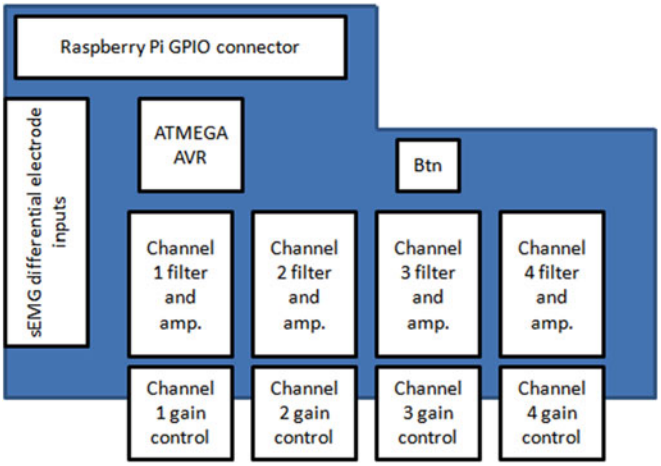

Raspberry Pi GPIO connector
sEMG differential electrode inputs
ATMEGA AVR
Btn
Channel 1 filter and amp.
Channel 2 filter and amp.
Channel 3 filter and amp.
Channel 4 filter and amp.
Channel 1 gain control
Channel 2 gain control
Channel 3 gain control
Channel 4 gain control

Fig. 3. Basic diagram of the custom PCB for the ECHO-OREAD device.

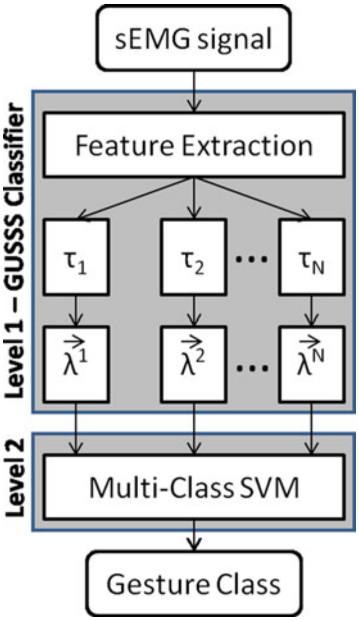


Fig. 4. Framework for the HiGUSSS Classifier used.

1) Hardware: The hardware of the device consists of a Raspberry Pi board with a custom-built PCB docked on top. The custom-built PCB contains circuitry for analog to digital conversion and four channels of sEMG inputs. The circuit provides amplification and individual, manual control of the gains for each channel. The channels are also filtered in order to reject undesirable frequencies. Once the signals are amplified and filtered, they are digitized and transferred to the Raspberry Pi through its GPIO connector. Additional buttons on the top of the device can be used to control the behavior of the boards, such as resetting the acquisition and reinitializing the boards. Fig. 3 shows a basic diagram of the custom PCB built for the ECHO-OREAD device.

III. Proposed Method

This research expands on the classification approach presented in [16] to further demonstrate the validity of performing sEMG classification based on extralaryngeal muscle activity in the anterior neck, which underlies voice production for speech and nonspeech behaviors (voiced and unvoiced sounds, throat clear, swallowing, etc.). A major difference between this approach and the one in [32] is that four sEMG channels were used instead of just one. The proposed framework is illustrated in Fig. 4 and it consists of a two-level hierarchical classifier. At the first level, there is a set of original GUSSS-based classifiers; and in the final level of the hierarchy a multiclass support vector machine (MC-SVM) performs the classification based on the outputs from the GUSSS classifiers. Basically, the GUSSS classifiers function as confidence generators, inputting feature vectors extracted from the raw sEMG signal and outputting N confidence vectors $\vec{\lambda}$, where the elements of the vector indicate the confidence that a crosstalk sEMG signal contains one of the sought-out signatures in the tuple—a tuple is a group with an arbitrary number of signatures: e.g., doubles, triples, etc. All of the obtained confidence vectors are concatenated into a second feature vector, which is then input to the SVM classifier at the second level of the hierarchy. The output of the second level classifier is the final class assigned to the observed sEMG signal. The following sections describe in further detail the classifiers at each level, as well as their training process.

A. Class Signatures

Let us assume that there is a labeled training set with $C \times T$ signals—i.e., T signals from each of the C possible classes (muscle patterns or gestures). First, a signature for each class

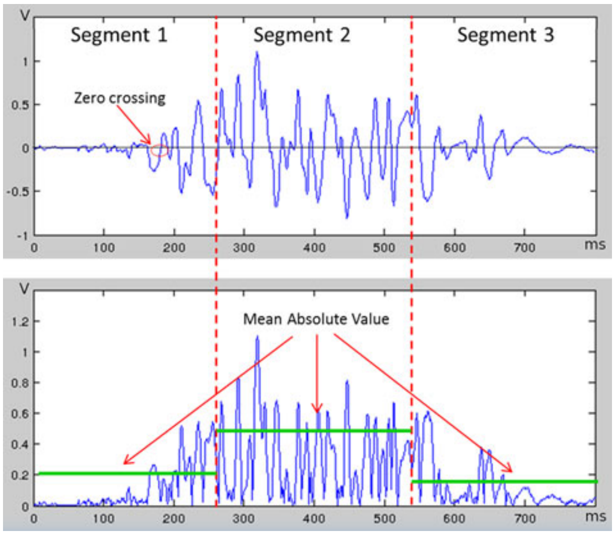


Fig. 5. Typical sEMG signal segmented into three parts. The ZCs are indicated in the top figure. The rectified signal and the MAVs of the segments are shown in the bottom figure.

is obtained. The current approach is to do an averaging of the T training signals grouped per class. That is, for each class c, a single signature: $s_c = \frac{1}{T}(\sum_{x_l \text{ in class } c} x_l)$ is created, where x_l is the lth training signal of class c.

B. Optimal Choice of Tuples

Each GUSSS-based classifier is associated to a tuple of classes, where the sizes and members can be chosen arbitrarily depending on the gestures, user, muscle activity patterns, etc. The rationale behind the tuples is the following: when a large number of C classes are considered at the same time, there might be much confusion between some of the classes. However, it is possible to find subsets of classes for which the confusion between such classes is minimized. So, the goal of the tuples is to allow similar classes to be separated. In a real-time system, it is desirable to group as many classes as possible per tuple in order to reduce the complexity of the algorithm, however, for this paper, pairs were chosen for the tuple numbers in order to achieve the highest accuracy possible.

The optimal pairings of gestures were automatically selected by iterating through all possible combinations and performing a modified version of a within class and between class analysis [33]—in the Euclidean space instead of the covariance space. The underlying equations and further details of this analysis can be found in [34].

C. sEMG Segmentation and Level 1 Feature Vectors

As mentioned before, the input to each of the GUSSS-based classifiers is a feature vector extracted from the incoming sEMG signal. The features used and the way to obtain the feature vector for a particular tuple, denoted τ_i, is described next. A same procedure applies to all N tuples being considered. Fig. 5 depicts a typical sEMG signal and the features considered.

1) GUSSS Ratio: As explained in Section II, the main idea of the GUSSS method is to identify particular signatures within a measured sEMG signal. For any given sEMG signal x, the GUSSS method seeks to identify the presence or not of each possible signatures. This is done by iteratively injecting signatures and obtaining ratios for each one of them. For all $n_i = |\tau_i|$ classes in tuple τ_i, the algorithm obtains the ratios $r_1, \ldots, r_{n_i}$. If signal x contains a pattern in class c, ratio r_c is expected to be smaller than all other ratios r_j, for $j \neq c$.

2) Segmentation of the sEMG Signals: Typically, the sEMG signals for the gestures considered here last from around 250 to 500 ms. To capture the structural information of the sEMG signals, we divide them into D segments of equal length. The features described next are calculated for each segment of any given signal.

3) Mean Absolute Value (MAV): One feature commonly used for sEMG signals is the MAV. The MAV of a signal $x(t)$ is obtained by calculating the average of the absolute values of x at all instants t. For a discrete signal:

$$\text{MAV} = \frac{1}{K}\sum_{k=1}^{K} |x(k)| \tag{1}$$

where K is the number of samples in a segment of x.

4) Zero Crossing (ZC): Another feature extracted from the sEMG signals is the number of ZC, which represents how many transitions from positive to negative (or vice-versa) there are in a segment of the signal.

5) Complete Feature Vector Level 1: After all of the features described above have been extracted, signal x is represented by the following feature vector:

$$\vec{v}_i = [r_1, \ldots, r_{n_i}, m_1, \ldots, m_D, z_1, \ldots, z_D] \tag{2}$$

where $r_1, \ldots, r_{n_i}$ are the GUSSS ratios for each class in tuple τ_i. The MAVs and ZCs for each segment of the signal are m_k and z_k, respectively, for $k = 1, \ldots, D$.

6) Statistics in Each Tuple of Gestures: As it will be shown shortly, the system uses the mean vector and covariance matrix of each class within the tuples. So, the above feature vectors are extracted for all T training signals in each class and used to form $\aleph(\vec{\mu}_j^i, \sum_j^i)$, representing the distribution of class j in the tuple τ_i, where $j = 1, \ldots, n_i$, and $i = 1, \ldots, N$.

D. Distances and Confidence Values

As it was mentioned before, the output of the first level in the hierarchy is a set of confidences that are concatenated to form a second feature vector for the next level. These confidences, which are based on Mahalanobis distances, are obtained by each one of the GUSSS-based classifiers.

First, an input signal y is fed into each one of the tuples described above. Then, for each tuple τ_i, a feature vector $\vec{v}_i$ (2) is calculated. Finally, the GUSSS-based classifiers calculate Mahalanobis distances to the mean vectors $\vec{\mu}_j^i$ of the classes in tuple τ_i, that is:

$$d_j^i = \sqrt{\left(\vec{v}_i - \vec{\mu}_j^i\right)\left(\sum\nolimits_j^i\right)^{-1}\left(\vec{v}_i - \vec{\mu}_j^i\right)^T}, \; j = 1, \ldots, n_i. \tag{3}$$

If, for example, distance d_j^i is small (close to zero), the confidence that signal y belongs to class j would be high.

To obtain the actual confidence values, the complementary error function is used:

$$\lambda(d_j^i) = \operatorname{erfc}\left(\frac{d_j^i}{\sqrt{2}}\right) \tag{4}$$

where $\operatorname{erfc}(x) = 1 - \operatorname{erf}(x)$.

For the GUSSS-based classifier corresponding to tuple τ_i, the confidence that signal y belongs to class j is given by $\lambda_j^i = \lambda(d_j^i)$. In the end, the classifier produces n_i confidence levels: $\vec{\lambda}^i = (\lambda_1^i, \ldots, \lambda_{n_i}^i)$.

Level 2 Feature Vector After confidence values are obtained for all N tuples, the second feature vector is created as follows:

$$\vec{u} = \left[\vec{\lambda}^1, \vec{\lambda}^2, \ldots, \vec{\lambda}^N\right]. \tag{5}$$

E. Multichannel HiGUSSS

For the enhanced version of the HiGUSSS with multiple channels used in this research, the steps detailed above are replicated for each channel, leading to a set of vectors $\overrightarrow{u}$. These channel vectors are then averaged in order to form a single confidence feature vector to serve as the input to the MC-SVM.

F. Level 2 Classifier: MC-SVM

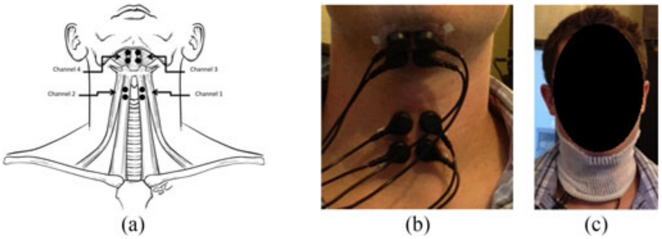
The final classification method consists of an MC-SVM. To train the MC-SVM, the$\overrightarrow{u}$ vectors are computed for all training signals, for all classes. When it comes to classification, an incoming signal $\overrightarrow{y}$ is fed through level 1 in the hierarchy to obtain the confidences and to create the $\overrightarrow{u_y}$ feature vector. The latter is fed into the MC-SVM in order to generate the final class assignment.

IV. Experiments

In this section, we address the research questions posed in Section I. That is, first to verify whether sEMG devices can reliably associate a larger number of sEMG patterns to speech and nonspeech behaviors; and second whether they can differentiate between multiple vowel sounds produced in a normal compared to a pressed manner. Also, with respect to normal and pressed vocal gestures, intra and intersubject testing was performed.

Expanding on the six gestures collected in [16] for a single subject, for the experiments reported here, a total of ten vocal gestures and one resting condition were collected for ten subjects. The full set of gestures collected for each subject includes: */a/, /a/ pressed, /u/, /u/ pressed, /i/, /i/ pressed, /t/, /s/, cough, and throat clear*. A pressed gesture indicates a simulated vocal dysfunction by using a pressed voice for the corresponding gesture as described in Section II. Vocally healthy subjects were trained on how to simulate these dysfunctional gestures by using a pressed voice, or restricting their air flow during a vocal gesture. This training is described in Section IV-A2.

These ten gestures were then grouped into the following four different test sets. The first test set includes the six original vocal gestures */u/, /i/, /t/, /s/, cough, and throat clear* also found in [16], but this time for ten subjects instead of one. The results for this test set will be discussed in Section V-A. The second test set consists of all ten gestures */a/, /a/ pressed, /u/, /u/ pressed, /i/, /i/ pressed, /t/, /s/, cough, and throat clear*. This set was used to test the ability of the hierarchical approach to classify a large gesture set with high accuracy and the results can be found in Section V-B. The next set was formed by the gestures */a/, /a/ pressed, /u/, /u/ pressed, /i/, and /i/ pressed*, and it was used to test the ability of the system to classify unique normal and simulated dysfunctional gestures. In other words, to identify a vocal gesture with or without simulated dysfunction, as well as to classify the occurrence of specific vocal gestures. The results for this test can be found in Section V-C. Finally, the gestures */a/, /u/, and /i/* were grouped together into a *Normal* class, while */a/ pressed, /u/ pressed, and /i/ pressed* into a *Pressed* class. The results for this test can be seen in Section V-D. In order to further validate the proposed hierarchical approach, a comparison found in Section V-E was performed between the proposed method, a distance classifier, and a single-layer MC-SVM classifier.



[redacted]

Fig. 6. Muscle groups on the human neck: diagram, actual view of electrode placement, and actual view with bandage applied.

A. Data Collection

The main goal of all experiments was to validate the claim that meaningful classification can be achieved from extralarygneal sEMG signals of the anterior neck—not only for normal voice production, but also simulated disordered voice production. Therefore, sEMG signals were collected under well-controlled laboratory conditions. The subjects were six males and four females in good health who denied the presence of any voice disorder. Four pairs of sEMG electrodes and a ground electrode were placed as explained in Section IV-A1 and seen in Fig. 6. Data were collected in an IAC Acoustics audiology booth (New York, New York) in the Department of Communication Science and Disorders, University of Missouri.

For the sEMG data collection, the test subjects were asked to perform 55 repetitions of each of the ten selected gestures in the following order: */a/, /a/ pressed, /u/, /u/ pressed, /i/, /i/ pressed, /t/, /s/, cough, and throat clear*. Each subject performed all of the gestures of a given type within a 2 s interval per each repetition of a gesture and with a 1 s break between repetitions. After the data for a single gesture were collected, the subject rested for several seconds and drank water as needed before completing the data collection for the next gesture.

During the rest period between repetitions of a gesture, the subject was asked to be as relaxed as possible, and try to minimize any motion in the throat or mouth area. The sEMG signals of interest, i.e., the ones to be associated with each gesture, are

TABLE I
MEASUREMENTS FOR EACH OF THE PARTICIPANTS' AIR FLOWS IN L/S FOR THE SET OF SIX NORMAL AND SIMULATED PRESSED VOICE GESTURES

	/a/		/u/		/i/	
Air Flow (L/s)	Normal	Pressed	Normal	Pressed	Normal	Pressed
Subject 1 (male, 40–59)	0.13	0.02	0.07	0.01	0.07	0.01
Subject 2 (female, 18–39)	0.14	0.02	0.15	0.04	0.17	0.04
Subject 3 (male, 18–39)	0.21	0.08	0.21	0.11	0.17	0.11
Subject 4 (male, 40–59)	0.17	0.07	0.26	0.10	0.18	0.08
Subject 5 (male, 18–39)	0.10	0.05	0.08	0.05	0.09	0.07
Subject 6 (female, 18–39)	0.11	0.07	0.11	0.06	0.10	0.06
Subject 7 (male, 18–39)	0.23	0.14	0.27	0.11	0.19	0.12
Subject 8 (male, 18–39)	0.16	0.13	0.20	0.18	0.19	0.15
Subject 9 (female, 18–39)	0.18	0.08	0.20	0.10	0.14	0.09
Subject 10 (female, 18–39)	0.12	0.03	0.15	0.05	0.12	0.06

those generated during the transition from the resting condition to the actual vocal gesture and back to resting.

For the experiments presented here, data were collected using a Tektronix MSO 4054 digital oscilloscope with a sample rate of 5 KHz. The signals were treated by both digital and analog bandpass filters at 30 Hz and 1 KHz and were divided into three segments (i.e., $D = 3$), as described in Section III-C2.

1) Electrode Placement: As seen in Fig. 6, surface electrodes were placed according to established guidelines for sEMG recordings [35] with special consideration of recommendations proposed for voice, speech, and swallowing research [6]. Disposable 10 mm Ag/AgCl surface electrodes (Bio-Medical Instruments, Warren, MI) were placed in bipolar configurations for single differential recordings from the anterior neck musculature. Two identical electrode pairs were placed on the left and right side of the neck to capture suprahyoid (submental) and infrahyoid muscular activity corresponding to elevations and depressions of the larynx during voice for speech, respectively [36]. The first electrode for the submental muscle site was placed approximately 1 cm from midline in the submandibular area superior to the hyoid bone [19], [37]–[39]. The second electrode of the submental pair was placed in line with the fibers of the muscle and with an interelectrode distance of approximately 1.5 cm [5], [6], [35], [40]. The submental location captures muscle activity from the anterior belly of the digastric, mylohyoid, and geniohyoid muscles.

For the infrahyoid muscle site, the first electrode was centered over the thyroid cartilage just below the thyroid notch and approximately 1 cm off midline [5], [6], [19], [37], [41]. The infrahyoid location captures muscle activity from the sternohyoid and omohyoid muscles with additional activity captured from the thin muscle sheath called platysma overlying most of the neck [6], [38]. Due to the small sizes of the individual muscles making up the submental and infrahyoid musculature as well as the multilayered structure of the muscles, sEMG can only capture muscle group activity and not activity from individual muscles. Moreover, it is not realistic to record activity from deeper muscles such as the thyrohyoid and cricothyroid [6]. The ground electrode was placed on the superior bony prominence of the shoulder (acromion). For voice and speech sEMG recordings, a placement of the ground electrode close to the electrodes is preferred [6]. A net bandage was placed over the electrodes in order to help keep the cables from moving around during data collection as seen in Fig. 6.

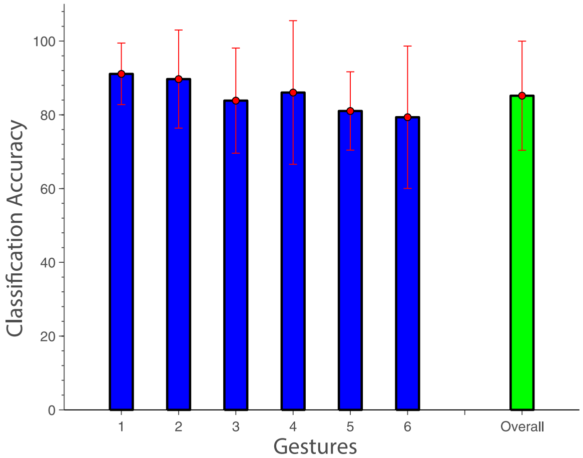


Fig. 7. Means and standard deviations of the classification accuracies per gesture, over all ten subjects. Six gestures considered: */u/, /i/, /t/, /s/*, *cough*, and *throat clear*.

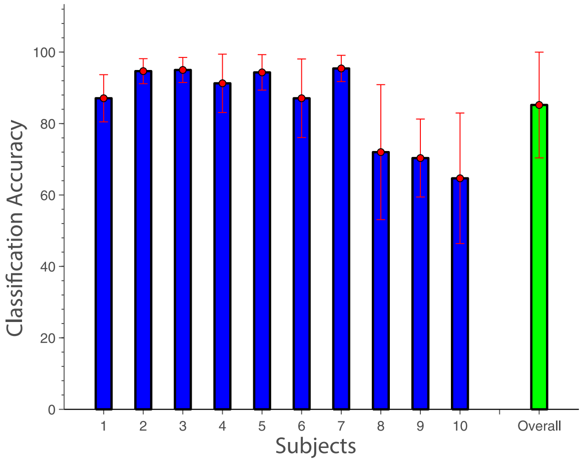


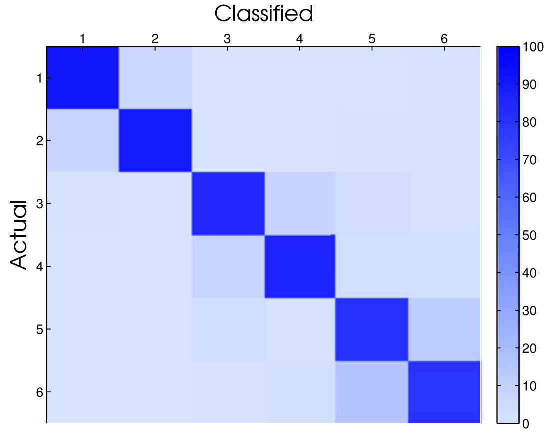
Fig. 8. Means and standard deviations of the classification accuracies per subject, over six gestures. Six gestures considered: */u/, /i/, /t/, /s/, cough,* and *throat clear*.

The quality of electrode placement was confirmed with tasks that produce target activations such as a swallow (submental and infrahyoid activity) and production of a front vowel (/i/, submental) and back vowel (/u/, infrahyoid).

2) Pressed Vocal Gesture Training: Since the participants were all vocally healthy, prior to data collection, a training program was implemented by a certified speech-language pathologist with experience in voice disorders to simulate "pressed" voice productions to be completed by each subject. The training consisted of verbal description and demonstrations of pressed voice based on the protocol by [24]: i.e., "an extremely high-effort phonation mode, with the perception of an almost completely closed airway, as if pushing". Next, subjects listened to selected audio samples of sustained /a/ productions by males and females with severe vocal hyperfunction chosen from the KayPentax Disordered Voice Database (Model 4337, Lincoln Park, NJ). Finally, the participants practiced normal

TABLE II
CONFUSION MATRIX FOR SIX GESTURES AVERAGED OVER ALL TEN SUBJECTS AND PRESENTED IN NUMERICAL AND GRAPHICAL FORMS

		Classified					
		1) */u/*	2) */i/*	3) */t/*	4) */s/*	5) *Cough*	6) *Throat clear*
Actual	1) */u/*	91.10	6.30	0.25	0.70	0.70	0.95
	2) */i/*	7.70	89.70	0.20	0.20	0.80	1.40
	3) */t/*	1.60	0.20	83.85	8.90	3.80	1.65
	4) */s/*	0.25	0.40	8.05	86.05	2.65	2.60
	5) *Cough*	0.70	0.60	3.95	1.45	81.05	12.25
	6) *Throat Clear*	0.25	0.60	1.25	2.65	15.90	79.35



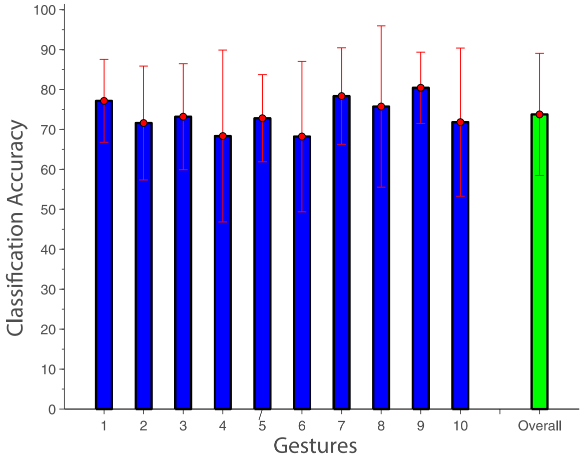


Fig. 9. Means and standard deviations of the classification accuracies per gesture, over all ten subjects. Ten gestures considered: */a/, /a/ pressed, /u/, /u/ pressed, /i/, /i/ pressed, /t/, /s/, cough,* and *throat clear*.

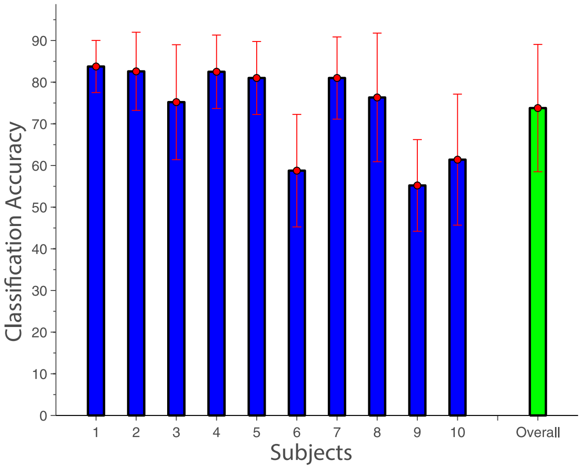


Fig. 10. Means and standard deviations of the classification accuracies per subject, over all ten gestures. Ten gestures considered: */a/, /a/ pressed, /u/, /u/ pressed, /i/, /i/ pressed, /t/, /s/, cough,* and *throat clear*.

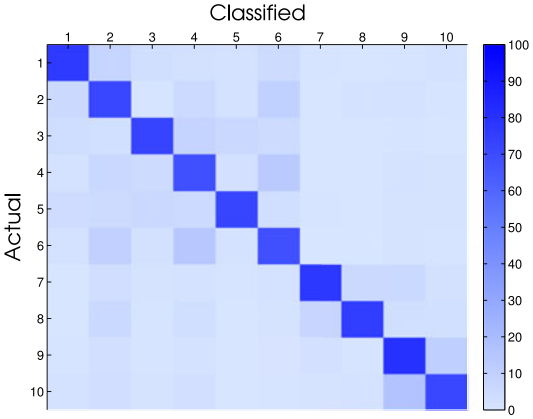
and pressed productions of /a/, /u/, and /i/ before proceeding to vowel productions with concurrent air-flow feedback using the KayPentax Phonatory Aerodynamic System. Target levels for air flow were primarily based on normative data for the PAS “Comfortable Sustained Phonation Protocol” (sustained /a/) and consisted of a mean of 0.13 L/s (SD = 0.08) (for both male and female, ages 18–39 years) and a mean of 0.11 L/s (SD = 0.05) (for males, 40–59 years) for normal /a/ production [42]. Reported normative data for average air-flow rates for the vowel /i/ are 0.14 L/s for males and 0.18 L/s for females [43]. Normative data for /u/ are not readily available, but they are expected to fall within a similar range. Each subject was able to produce normal and pressed vowel productions inside and outside the norm range, respectively, during training, and contrasts were perceptually distinct. Five repetitions of each vowel gesture were recorded with concurrent air-flow visual feedback data prior to full sEMG data collection. Full sEMG data were collected without concurrent air-flow feedback to avoid additional muscular neck activity from holding the PAS face mask against the face. During data collection, all participants were perceptually monitored for contrasts between normal and pressed productions. Participants were encouraged and received feedback to maintain pressed phonations as necessary. The average air flows over each gesture are presented in Table I.

V. RESULTS

For all of the results presented here, a ten-fold cross validation was performed. Each time 90% of the signals from all collected

TABLE III
CONFUSION MATRIX FOR TEN GESTURES AVERAGED OVER ALL TEN SUBJECTS AND PRESENTED IN NUMERICAL AND GRAPHICAL FORMS

		Classified									
		1) */a/*	2) */a/ pressed*	3) */u/*	4) */u/ pressed*	5) */i/*	6) */i/ pressed*	7) */t/*	8) */s/*	9) *Cough*	10) *Throat Clear*
Actual	1) */a/*	77.17	7.90	3.08	2.20	1.90	4.70	0.40	0.25	0.75	1.65
	2) */a/ pressed*	5.78	71.62	0.58	5.55	1.98	10.08	0.25	1.55	2.00	0.60
	3) */u/*	4.10	2.65	73.20	8.27	6.45	4.88	0.00	0.00	0.45	0.00
	4) */u/ pressed*	1.45	6.65	4.70	68.35	2.30	13.25	0.20	0.25	1.85	1.00
	5) */i/*	4.20	4.95	6.25	5.55	72.82	3.68	0.40	0.20	1.05	0.90
	6) */i/ pressed*	2.00	10.10	2.28	14.05	1.35	68.22	0.20	0.25	1.10	0.45
	7) */t/*	0.00	3.30	1.00	1.65	0.40	1.32	78.37	5.77	6.00	2.20
	8) */s/*	0.00	6.10	0.50	3.45	0.00	0.83	7.77	75.75	3.00	2.60
	9) *Cough*	0.65	2.80	0.50	1.50	0.00	0.33	2.40	0.53	80.43	10.85
	10) *Throat Clear*	1.60	3.25	1.00	2.90	0.20	0.45	0.85	1.40	16.53	71.82



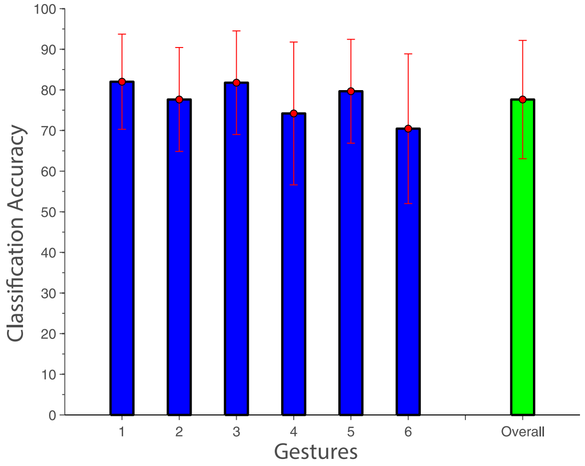


Fig. 11. Means and standard deviations of the classification accuracies per gesture, over all subjects. Six gestures considered: */a/, /a/ pressed, /u/, /u/ pressed, /i/,*and */i/ pressed*.

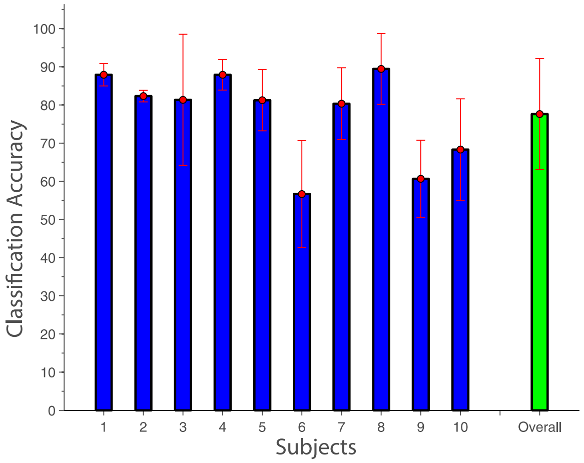


Fig. 12. Means and standard deviations of the classification accuracies per subject, over six gestures. Six gestures considered: */a/, /a/ pressed, /u/, /u/ pressed, /i/,*and */i/ pressed*.

gestures were used for training and the remaining 10% were used for classification.

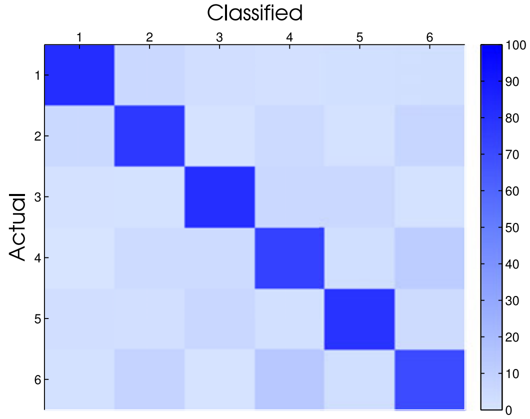
A. *Distinct Gestures*

In the first test, a set of six distinct gestures containing speech and nonspeech behaviors (e.g., vowel, consonant, and throat sounds) was used. These are the same gestures used in our previous work [16], that is, */u/, /i/, /t/, /s/, cough, throat clear*. However, the results here expand upon those tests by using data from ten subjects—four female and six males—as opposed to a single subject. Classification was completed using the improved HiGUSSS algorithm described in Section III. The overall average in classification accuracy—i.e., over all gestures and over all subjects—was approximately 85%. Fig. 7 shows the classification accuracies per gesture, averaged over all ten subjects, while Fig. 8 shows the classification accuracies per subject, averaged over all six gestures. Both the means and the corresponding standard deviations are depicted in these same figures. Finally, Table II presents the average confusion matrix computed over all ten subjects. A color plot of the same Table II is provided next to its numerical form for better visualization of the results.

The high accuracy achieved in this test should positively address part of our first research question: whether sEMG devices

TABLE IV
CONFUSION MATRIX FOR THE SECOND SET OF SIX GESTURES AVERAGED OVER ALL TEN SUBJECTS AND PRESENTED IN NUMERICAL AND GRAPHICAL FORMS

		Classified					
		1) */a/*	2) */a/ pressed*	3) */u/*	4) */u/ pressed*	5) */i/*	6) */i/ pressed*
Actual	1) */a/*	82.00	6.38	3.33	2.45	2.65	3.18
	2) */a/ pressed*	6.15	77.63	1.30	5.40	2.00	7.52
	3) */u/*	1.83	1.70	81.77	6.33	6.30	2.07
	4) */u/ pressed*	0.65	4.75	4.90	74.20	3.65	11.85
	5) */i/*	3.30	2.90	6.70	2.65	79.67	4.78
	6) */i/ pressed*	2.10	8.65	1.00	14.20	3.60	70.45



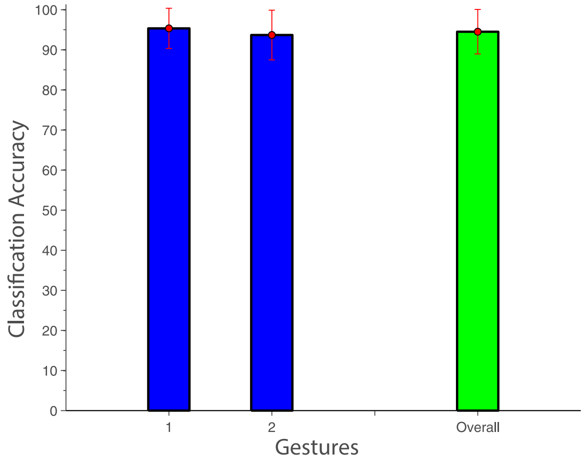


Fig. 13. Means and standard deviations of the classification accuracies over all subjects using normal versus pressed gestures—i.e., simulated dysfunction.

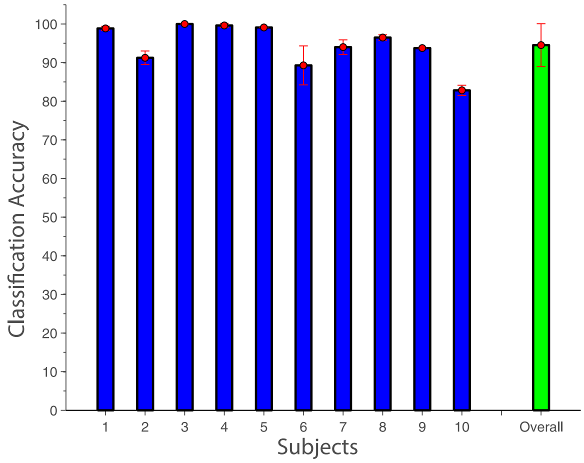
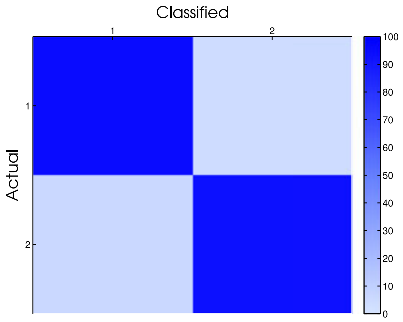


Fig. 14. Means and standard deviations of the classification accuracies per subject using normal versus pressed gestures—i.e., simulated dysfunction.

TABLE V
CONFUSION MATRIX FOR INTRASUBJECT ACCURACY IN DETECTION OF SIMULATED DYSFUNCTION AVERAGED OVER ALL TEN SUBJECTS AND PRESENTED IN NUMERICAL AND GRAPHICAL FORMS

		Classified	
		1) Normal	2) Pressed
Actual	1) Normal	95.35	4.65
	2) Pressed	6.32	93.68



can reliably associate a larger number of patterns of extralaryngeal muscle activity with voice tasks underlying speech and nonspeech behaviors.

B. *Large Gesture Set*

In order to further address our first research question, we selected a large number of gestures (ten). The selected gestures were, */a/, /a/ pressed, /u/, /u/ pressed, /i/, /i/ pressed, /t/, /s/, cough, throat clear*. As earlier, classification was completed using the improved HiGUSSS algorithm described in Section III. An overall average classification accuracy of 74% was achieved—i.e., averaged over all gestures and over all subjects. As it can be observed in the next figures, it is important to notice that one of the subjects (#6) presented a much lower average, bringing down the overall average to 74%. In spite of that, these results still shows that the system is robust to large

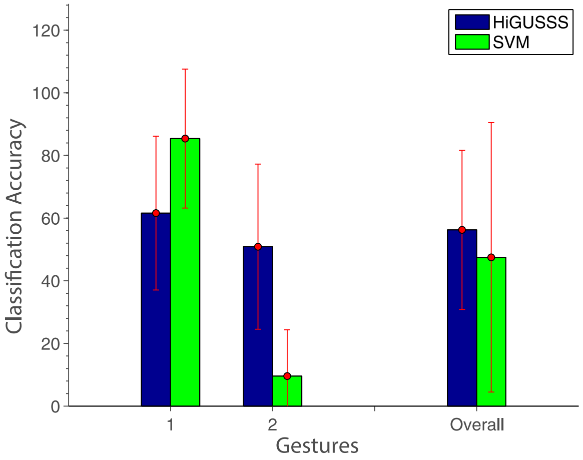


Fig. 15. Means and standard deviations of the classification accuracies per gesture, for normal versus pressed gestures—using the leave-one-out approach for the HiGUSSS and the SVM Classifier.

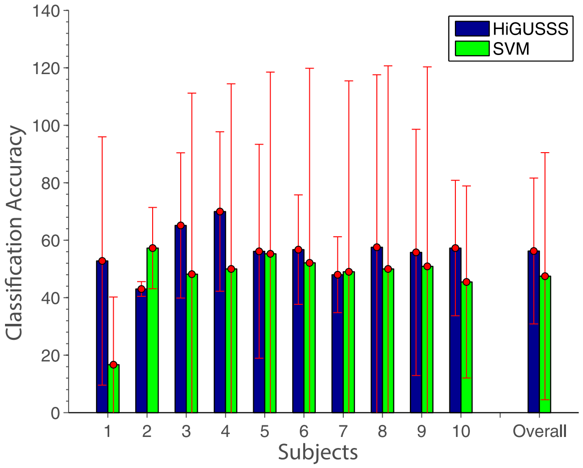
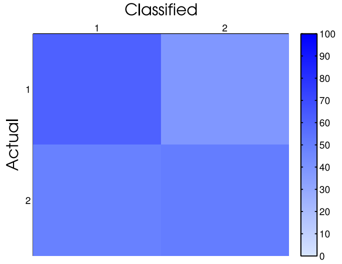


Fig. 16. Means and standard deviations of the classification accuracies per subject, for normal versus pressed gestures—using the leave-one-out approach for the HiGUSSS and the SVM Classifier.

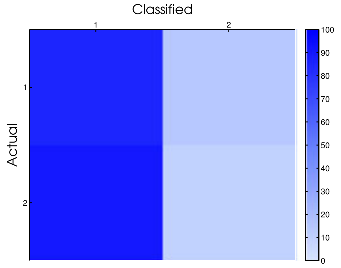
TABLE VI

Confusion Matrix for Intersubject Accuracy in Detection of Simulated Dysfunction Averaged Over All Ten Subjects and Presented in Numerical and Graphical Forms for the HiGUSSS and the SVM Classifier

HiGUSSS		Classified	
		1) Normal	2) Pressed
Actual	1) Normal	61.60	38.40
	2) Pressed	49.12	50.88



SVM		Classified	
		1) Normal	2) Pressed
Actual	1) Normal	85.40	14.60
	2) Pressed	90.42	9.58



gesture sets (research question one), but also to sets containing similar gestures (research question two). Figs. 9 and 10 show, respectively, the classification accuracies per gesture, over all ten subjects, and per subject, over all ten gestures. As in the previous test, Table III shows the average confusion matrix in both numerical and graphical forms.

C. Normal versus Pressed Gestures

In order to show that not only can the system achieve high accuracy with similar gestures, but it can also distinguish specific gestures within the normal and pressed classes, tests were completed using three normal and three pressed gestures. This partially addresses our second research question or whether the system can differentiate between multiple vowel sounds produced in a normal manner compared with a pressed (low air flow) manner.

The gestures used here were */a/, /a/ pressed, /u/, /u/ pressed, /i/,*and */i/ pressed*, and they were collected as described in Section IV-A2. In this case, an overall average classification accuracy of 78% was achieved. As before, classification accuracies per gesture and per subject are presented before the average confusion matrix: Figs. 11, 12, and Table IV, respectively.

D. Intra and InterSubject Testing

The remainder of research question two was addressed by completing intra and intersubject testing. Once again, together with the previous test, the results of this test can justify our method as a potential solution to the detection of normal and maladaptive extralaryngeal patterns associated with voice problems.

1) Intrasubject: Intrasubject testing was carried out for each of the ten subjects separately using three normal and three pressed gestures, that is, */a/, /a/ pressed, /u/, /u/ pressed, /i/,* and */i/ pressed*. These gestures were divided into two sub-classes: the gestures */a/, /u/, and /i/* were treated as the *Normal* class, and gestures */a/ pressed, /u/ pressed, and /i/ pressed* were treated as the *Pressed* class. The classification was then completed as a two class problem using the method described in Section III. An overall average in classification accuracy of 95% was achieved. This clearly support the potential use of an sEMG device for detecting vocal dysfunctions. Fig. 13 shows the average classification accuracy over all subjects for each of the two sub-classes (Normal versus Pressed), and Fig. 14 shows the results per user. Table V shows the average confusion matrix averaged over all ten subjects.

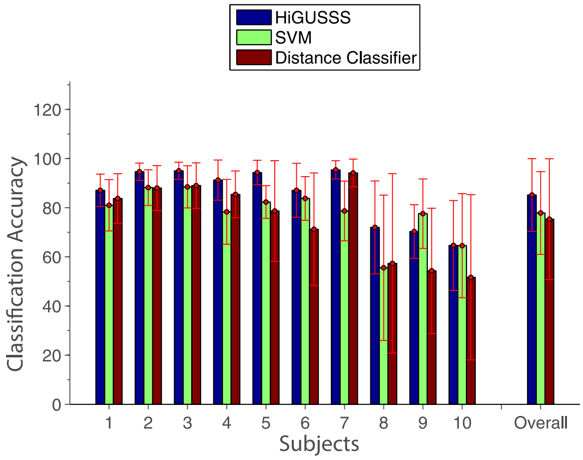


Fig. 17. Comparison between the average classification accuracies per subject, over six gestures, for each of the three classifier (HiGUSSS, MC-SVM, and Distance). First set of six gestures considered: */u/, /i/, /t/, /s/, cough,* and *throat clear*.

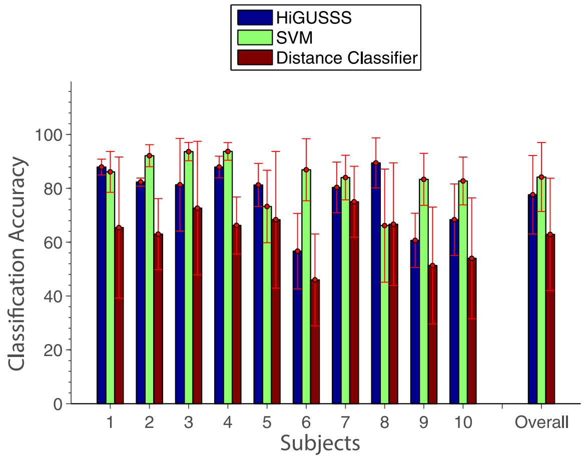


Fig. 18. Comparison between the average classification accuracies per subject, over six gestures, for each of the three classifier (HiGUSSS, MC-SVM, and Distance). Second set of six gestures considered: */a/, /a/ pressed, /u/, /u/ pressed, /i/,* and */i/ pressed*.

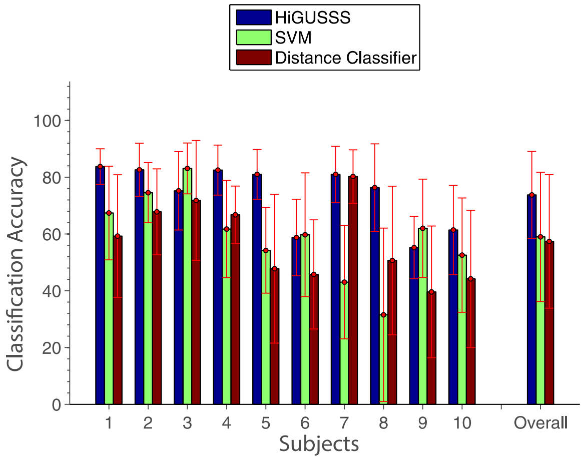


Fig. 19. Comparison between the average classification accuracies per subject, over all ten gestures, for each of the three classifier (HiGUSSS, MC-SVM, and Distance). Set of all ten gestures considered: */a/, /a/ pressed, /u/, /u/ pressed, /i/, /i/ pressed, /t/, /s/, cough,* and *throat clear*.

2) Intersubject: Intersubject tests were also completed for all ten subjects combined and using the same three normal and three pressed gestures. These gestures were once again grouped into two sub-classes, *Normal* and *Pressed*. However, this time, the test was performed in a *leave-one-out*fashion. That is, for each subject, the training was completed using data from all the other subjects—i.e., all data except for the subject being tested. Classification was then completed as a two class problem using the data from that subject and the method described in Section III.

As before, Figs. 15 and 16, and Table VI show the results of this test. As the reader will notice, the HiGUSSS performed very poorly for the intersubject case. So, to further investigate the reason for such low performance, we run the same test set using an MC-SVM. The figures and tables above also include the result for the MC-SVM, which performed even worse than the HiGUSSS. We attribute this poor performance by both classifiers to data overfitting: i.e., obviously, both classifiers can learn very well each subject's patterns, but they fail to generalize across subjects. This conclusion is supported by the excellent result in the intrasubject test on one hand, and the poor result in inter-subject test on the other hand. The actual reason for this overfitting needs to be further investigated, but adding more diversity to the data by adding more subjects or more gestures could alleviate this problem.

E. Classifier Comparison

Finally, a comparison between the HiGUSSS algorithm and two other classifiers was also completed. The goal was to illustrate the value of the HiGUSSS system as opposed to other more traditional classifiers. For comparison purposes, tests were run on all three groups of gestures: i.e., the six distinct gestures used in Section V-A; the large gesture set used in Section V-B; and the three pressed and three normal gestures used in Section V-C. These three groups of gestures were classified using both a simple distance classifier and a single layer MC-SVM, which were then compared to the results for the HiGUSSS presented earlier. Figs. 17, 18, and 19 show the corresponding classification results per subject, and overall. Note that the HiGUSSS classifier outperforms the distance and SVM classifiers in almost all the cases, and most notably for the cases with large number of gestures.

VI. Discussion and Conclusion

The results presented in Section V-A for the distinct gestures further reinforce conclusions drawn from previous study [16], and address our first research question: that meaningful classification can be drawn from sEMG signals collected at the anterior neck. In [16], data were collected and tested for only a single subject, while in this study data were collected and tested for ten subjects (four females and six males). As seen in Section V-A, average classification accuracy for ten subjects performing six gestures (*/u/, /i/, /t/, /s/, cough, throat clear*) was 85%, which is consistent with the 90% classification accuracy achieved in [16] for one single subject.

The tests in Section V-B combined the gestures from Sections V-A and V-C in order to partially address both our first and second research questions. This gesture set allowed for the testing of the system for a larger number of gestures. The HiGUSSS system achieved an average classification accuracy of 74% for this combined set of ten gestures. This demonstrates the merits of the HiGUSSS system when it comes to larger gesture sets. As expected, the more distinguishable gestures from Section V-A (*/t/, /s/, cough, throat clear*) achieved a higher average classification accuracy than the set of gestures in Section V-C (*/a/, /a/ pressed, /u/, /u/ pressed, /i/, /i/ pressed*). This can be seen in the nondiagonal entries of Table III, which contains the average confusion matrix for the classification of the ten gestures.

Further addressing research question two, we strove to explore whether or not this system could classify, with high accuracy, unique normal gestures, and unique pressed gestures. This inspired the tests done in Section V-C where ten subjects performed both normal and simulated dysfunctional gestures. The HiGUSSS system achieved an average classification accuracy of 78%, over all ten subjects. Although six gestures were tested as in Section V-A, the drop in accuracy from 85% to 78% comes from the fact that the gestures tested in Section V-C were more similar to each other than the gestures tested in Section V-A. The gestures in Section V-A included vowel sounds, throat sounds, and consonants while the gestures in Section V-C included only vowel sounds as well as the repetition of same gesture performed in both a normal and pressed voice. Confusion occurred both between the corresponding normal and pressed gestures as well as within the set of pressed gestures and within the set of normal gestures.

After verifying that the system could detect unique normal and pressed vocal gestures in Section V-C, we concluded research question two in Section V-D where we tested the ability of the system to detect the presence of simulated dysfunctions for both inter and intrasubject conditions. As can be seen in Section V-D1, accuracy for detection of intrasubject simulated vocal dysfunction was 95%. This demonstrates the potential of applying a system like this to the early detection of vocal disorders through the detection of changing and/or emerging intrasubject patterns of sEMG signals.

Also, in order to further explore the potential of the proposed application of sEMG in detection of voice dysfunctions, intersubject recognition of simulated vocal dysfunction was tested. As expected, the average classification accuracy seen in Section V-D2 was lower than the accuracy achieved for intrasubject testing. This could be the result of the more unique character of vocal gestures for each subject. It is also possible that the training data were not diverse enough to allow the HiGUSSS to generalize the learned patterns, leading the classifier to overfit the data for each individual. In that case, by increasing the training set of vocal gestures with more test subjects could significantly improve the classification results. It is also important to note that the system could still be used with intersubject data in a fashion similar to many voice recognition systems that improve over time by continuously learning patterns from the current user.

Finally, the classification accuracy achieved using the HiGUSSS system was compared with the classification accuracies from a single-layer MC-SVM and a simple distance classifier in Section V-E. Two meaningful trends were discovered during these tests. First, it can be noticed that as the number of gestures increased, the advantage of the HiGUSSS method over the other two classifiers became clearer. Second, as the gestures in the set became more similar, once again the HiGUSSS system outperformed the other two methods. This shows the validity of the HiGUSSS system and the stronger case for its application in the detection of vocal dysfunctions for similar gestures and large vocal gesture sets.

Future work will focus on expanding the subject pool in order to collect more normative data and to test the system with data of patients with clinical vocal dysfunction as identified by vocal effort and vocal fatigue. Voice disorders occur on a continuum and it will be critical to correlate auditory-perceptual ratings of a pressed (strained) voice quality and altered phonatory aerodynamic function (air flow, air pressure, and laryngeal airway resistance) with sEMG data. Such data are routinely collected in clinical voice protocols based on vowel or consonant-vowel productions. This would allow for further exploration of the intra and intersubject testing presented here and refine the system's sensitivity and specificity to differentiate normal from dysfunctional voice productions. Next, the feasibility of using the HiGUSSS recognition system during ambulatory monitoring must be tested with occupational voice users, as for example, student teachers. In addition, expanding the gesture set will allow for a more accurate simulation of the usage case of this device. During normal speech countless vocal and nonvocal gestures occur (swallowing, coughing, consonants, vowels, etc.) with and without dysfunctions, so that the system must be able to classify very large gesture sets with high accuracy. Gestures may also include short target phrases that could be easily implemented during ambulatory monitoring as reference points. Overall, the HiGUSSS recognition system shows great promise in helping to better understand changes in vocal function that may be linked to voice disorders.

REFERENCES

[1] N. Roy, R. Merrill, S. Thibeault, R. Parsa, S. Gray, and E. Smith, "Prevalence of voice disorders in teachers and the general population," *J. Speech. Lang. Hear. Res.*, vol. 47, pp. 281–293, 2004.

[2] K. Verdolini, C. Rosen, and R. Branski, *Classification Manual for Voice Disorders*. Mahwah, NJ, USA: Lawrence Erlbaum Associates, 2006.

[3] S. Joshi, A. Wexler, C. Perez-Maldonado, and S. Vernon, "Brain-muscle-computer interface using a single surface electromyographic signal: Initial results," in *Proc. 5th Int. IEEE/EMBS Conf. Neural Eng.*, Apr. 2011, pp. 342–347.

[4] K. Singh, R. Bhatia, and H. Ryait, "Precise and accurate multifunctional prosthesis control based on fuzzy logic techniques," in *Proc. Int. Conf. Commun. Syst. Netw. Technol.*, Jun. 2011, pp. 188–193.

[5] C. Stepp, J. Heaton, M. Braden, T. Stadelman-Cohen, M. Jetté, and R. Hillman, "Comparison of neck tension palpation rating systems with surface electromyographic and acoustic measures in vocal hyperfunction," *J. Voice*, vol. 25, pp. 67–75, 2011.

[6] C. Stepp, "Surface electromyography for speech and swallowing systems: Measurement, analysis, and interpretation," *J. Speech. Lang. Hear. Res*, vol. 55, pp. 1232–1246, 2012.

[7] L. A. Rivera and G. N. DeSouza, *Haptic and Gesture-Based Assistive Technologies for People With Motor Disabilities*. Hershey, PA, USA: IGI Global, 2013, ch. 1, pp. 1–27.

[8] R. Hillman, J. Heaton, A. Masaki, S. Zeitels, and H. Cheyne, “Ambulatory monitoring of disordered voices,” *Ann. Otol. Rhinol. Laryngol.*, vol. 115, no. 11, pp. 795–801, 2006.

[9] P. Popolo, J. Švec, and I. Titze, “Adaptation of a pocket pc for use as a wearable voice dosimeter,” *J. Speech. Lang. Hear. Res.*, vol. 48, no. 4, pp. 780–791, 2005.

[10] J. H. van Stan, J. Gustafsson, E. Schalling, and R. E. Hillman, “Direct comparison of three commercially available devices for voice ambulatory monitoring and biofeedback,” *Perspectives on Voice and Voice Disorders*, vol. 24, no. 2, pp. 80–86, 2014.

[11] D. Mehta, M. Zañartu, S. Feng, H. Cheyne II, and R. Hillman, “Mobile voice health monitoring using a wearable accelerometer sensor and a smartphone platform,” *IEEE Trans. Biom. Eng.*, vol. 59, no. 11, pp. 3090–3096, Nov. 2012.

[12] R. E. Hillman, M. Zañartu, M. Ghassemi, D. D. Mehta, J. H. Van Stan, H. A. Cheyne II, and J. V. Guttag, “Future directions in the development of ambulatory monitoring for clinical voice assessment,” in *Proc. 10th Int. Conf. Adv. Quantitative Laryngol.*, 2013, pp. 23–24.

[13] D. D. Mehta, M. Zanartu, J. H. Van Stan, S. W. Feng, H. A. Cheyne II, and R. E. Hillman, “Smartphone-based detection of voice disorders by long-term monitoring of neck acceleration features,” in *Proc. IEEE Int. Conf. Body Sensor Netw.*, 2013, pp. 1–6.

[14] M. Zañartu, V. Espinoza, D. D. Mehta, J. H. Van Stan, H. A. Cheyne II, M. Ghassemi, J. V. Guttag, and R. E. Hillman, “Toward and objective aerodynamic assessment of vocal hyperfunction using a voice health monitor,” in *Proc. Models Anal. Vocal Emissions Biomed. Appl.: 8th Int. Workshop*, Firenze, Italy, 2013, pp. 167–170.

[15] J. L. Spielman, E. J. Hunter, A. E. Halpern, and I. R. Titze, “Measuring improvement in teachers with voice complaints using the inability to produce soft voice (IPSV): Preliminary data,” 2012.

[16] N. Smith, T. Klongtruagrok, G. DeSouza, C. Shyu, M. Dietrich, and M. Page, “Non-invasive ambulatory monitoring of complex sEMG patterns and its potential application in the detection of vocal dysfunctions,” in *Proc. 16th Int. Conf. e-Health Netw., Appl. Services*, Oct. 2014, pp. 447–452.

[17] R. E. Hillman, E. B. Holmberg, J. S. Perkell, M. Walsh, and C. Vaughan, “Objective assessment of vocal hyperfunctionan experimental framework and initial results,” *J. Speech, Lang., Hear. Res.*, vol. 32, no. 2, pp. 373–392, 1989.

[18] S. Y. Lowell, R. T. Kelley, R. H. Colton, P. B. Smith, and J. E. Portnoy, “Position of the hyoid and larynx in people with muscle tension dysphonia,” *Laryngoscope*, vol. 122, no. 2, pp. 370–377, 2012.

[19] M. Dietrich and K. V. Abbott, “Vocal function in introverts and extraverts during a psychological stress reactivity protocol,” *J. Speech, Lang., Hear. Res.*, vol. 55, no. 3, pp. 973–87, 2012.

[20] C. E. Stepp, R. E. Hillman, and J. T. Heaton, “Use of neck strap muscle intermuscular coherence as an indicator of vocal hyperfunction,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 18, no. 3, pp. 329–335, Jun. 2010.

[21] M. Dietrich and K. V. Abbott, “Psychobiological stress reactivity and personality in persons with high and low stressor-induced extralaryngeal reactivity,” *J. Speech, Lang., Hear. Res.*, vol. 57, pp. 2076–2089, Dec. 2014.

[22] N. Roy and D. M. Bless, “Toward a theory of the dispositional bases of functional dysphonia and vocal nodules: exploring the role of personality and emotional adjustment,” in *Handbook of Voice Quality Measurement*, R. Kent and M. Ball, Eds. San Diego, CA, USA: Singular Publishing Group, 2000.

[23] B. E. Kostyk and A. P. Rochet, “Laryngeal airway resistance in teachers with vocal fatigue: A preliminary study,” *J. Voice*, vol. 12, pp. 287–299, 1998.

[24] M. Dietrich and K. V. Abbott, “Evidence for distinguishing pressed, normal, resonant, and breathy voice qualities by laryngeal resistance and vocal efficiency in vocally trained subjects,” *J. Voice*, vol. 22, pp. 546–552, 2008.

[25] J. Stemple, B. Weinrich, and S. B. Brehm, “Aerodynamic measurement of vocal function: Phonatory aerodynamic system,” in *Handbook of Voice Assessments*. Plymouth, U.K.: Plural Publication, 2011, pp. 7–20.

[26] A. I. Gillespie, J. Gartner-Schmidt, E. N. Rubinstein, and K. V. Abbott, “Aerodynamic profiles of women with muscle tension dysphonia/aphonia,” *J. Speech, Lang., Hear. Res.*, vol. 56, pp. 481–488, Apr. 2013.

[27] E. U. Grillo, K. Perta, and L. Smith, “Aerodynamic profiles of women with muscle tension dysphonia/aphonia,” *Logopedics Phoniatrics Vocol.*, vol. 34, no. 1, pp. 43–48, 2009.

[28] L. A. Rivera and G. N. DeSouza, “Recognizing hand movements from a single sEMG sensor using guided under-determined source signal separation,” in *Proc. 12th IEEE Int. Conf. Rehabil. Robot.*, ETH Zurich, Switzerland, Jun. 2011, pp. 450–455.

[29] Y. Guangying, “Study of myoelectric prostheses hand based on independent component analysis and fuzzy controller,” in *Proc. 8th Int. Conf. Electron. Meas. Instrum.*, Aug. 2007, pp. 1-174–1-178.

[30] J. G. Webster, Ed,. *Encyclopedia of Medical Devices and Instrumentation. Electromyography*. New York, NY, USA: Wiley, 2006.

[31] L. A. Rivera and G. N. DeSouza, “A power wheelchair controlled using hand gestures, a single sEMG sensor, and guided under-determined source signal separation,” in *Proc. 4th IEEE RAS & EMBS Int. Conf. Biomed. Robot.*, Rome, Italy, Jun. 2012, pp. 1535–1540.

[32] L. A. Rivera, N. R. Smith, and G. N. DeSouza, “High-accuracy recognition of muscle activation patterns using a single sEMG signal,” in *Proc. 5th IEEE RAS & EMBS Int. Conf. Biomed. Robot. Biomechatron.*, Sao Paulo, Brazil, Aug. 2014, pp. 579–584.

[33] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. New York, NY, USA: Academic Press, 2013.

[34] N. R. Smith, “A hierarchical framework for pattern recognition with application in source signal separation,” M.S. thesis, Univ. Missouri, Columbia, MO, USA, May 2015.

[35] A. Fridlund and J. Cacioppo, “Guidelines for human electromyographic research,” *Psychophysiology*, vol. 23, no. 5, pp. 567–589, 1986.

[36] T. Hixon, G. Weismer, and J. Hoit, *Preclinical Speech Science. Anatomy, Physiology, Acoustics, Perception*. 2nd ed. Plymouth, U.K.: Plural, 2013.

[37] R. Ding, C. Larson, J. Logemann, and A. Rademaker, “Surface electromyographic and electroglottographic studies in normal subjects under two swallow conditions: Normal and during the mendelsohn manuever,” *Dysphagia*, vol. 17, pp. 1–12, 2002.

[38] A. VanBoxtel, “Optimal signal bandwidth for the recording of surface EMG activity of facial, jaw, oral, and neck muscles,” *Psychophysiology*, vol. 38, pp. 22–34, 2001.

[39] E. Yiu, K. Verdolini, and L. Chow, “Electromyographic study of motor learning for a voice production task,” *J. Speech. Lang. Hear. Res*, vol. 48, no. 6, pp. 1254–1268, 2005.

[40] C. DeLuca, “The use of surface electromyography in biomechanics,” *J. Appl. Biomechanics*, vol. 13, no. 2, pp. 135–163, 1997.

[41] R. Colton, J. Casper, and R. Leonard, *Understanding Voice Problems: A Physiological Perspective for Diagnosis and Treatment*. 3 ed. Baltimore, MD, USA: Lippincott Williams and Wilkins, 2006.

[42] R. I. Zraick, L. Smith-Olinde, and L. L. Shotts, “Adult normative data for the kaypentax phonatory aerodynamic system model 6600,” *J. Voice*, vol. 26, pp. 164–176, 2012.

[43] P. Woo, R. Colton, and L. Shangold, “Phonatory airflow analysis in patients with laryngeal disease,” *Ann. Otol., Rhinol. Laryngol.*, vol. 96, pp. 549–555, 1987.

[44] M. van Mersbergen, C. Patrick, and L. Glaze, “Functional dysphonia during mental imagery: Testing the trait theory of voice disorders”, *J. Speech, Lang., Hear. Res.*, vol. 51, no. 6, pp. 1405–1423, Dec. 2008.

Nicholas R. Smith received the undergraduate and Master’s degrees from the University of Missouri, Columbia, MO, USA, in electrical engineering.

His Master’s research focused on machine learning and pattern recognition and his thesis, which focused on different applications of hierarchical pattern recognition, was entitled “A Hierarchical Framework for Pattern Recognition Using Source Signal Separation.” In addition to his thesis, his work has been published in several conferences including Healthcom and BioRob. He is currently with Texas Instruments in Dallas.

Luis A. Rivera (S'11) received the Graduate degree in electronics engineering from Del Valle University, Guatemala city, Guatemala, in 2006. He received a Fulbright scholarship, and received the M.S. degree in electrical engineering from the University of Missouri, Columbia, MO, USA, in 2011. He is currently working toward the Ph.D. degree at the University of Missouri, working at the Vision-Guided and Intelligent Robotics Lab.

He worked for the Departments of Mathematics and Physics, Del Valle University from 2007 to 2009. His research interests include the areas of machine learning, pattern recognition, and robotic assistive technology. He has worked on systems and interfaces for controlling assistive technology devices such as power wheelchairs, using surface EMG signals, head motion, etc. He has also worked on methods for detecting patterns in mixed signals, with applications in terahertz and assistive technology.

Maria Dietrich received the Diploma degree in Heilpädagogik from the Universität zu Köln, Cologne, Germany, in 2001, the M.A. degree in speech-language pathology from Kent State University, Kent, OH, USA, in 2003, and the Ph.D. degree in communication science and disorders from the University of Pittsburgh, Pittsburgh, PA, USA, in 2009.

She was a Postdoctoral Scholar with the Department of Rehabilitation Sciences, University of Kentucky, and is currently an Assistant Professor in the Department of Communication Science and Disorders, University of Missouri, Columbia, MO, USA.

Chi-Ren Shyu (S'89–M'99–SM'07) received the Ph.D. degree in electrical and computer engineering from Purdue University, West Lafayette, IN, USA.

Upon completing one year of postdoctoral training with Purdue, he joined the Department of Computer Engineering and Computer Science, University of Missouri (MU), Columbia, MO, USA, in October 2000. He is currently the Shumaker Endowed Professor and the Chairman of the Department of Electrical and Computer Engineering. He also heads the MU Informatics Institute where 43 core faculty members from 17 departments from MU support an interdisciplinary training and research program in bioinformatics, health informatics, and geoinformatics. His research interests include biomedical informatics, big data analytics, and visual knowledge reasoning.

Dr. Shyu received the US National Science Foundation Faculty Early Career Development (NFS CAREER) Award, the MU College of Engineering Faculty Research Award, the MU Faculty Entrepreneurial Award, and various teaching awards. He is a Member of the American Association for the Advancement of Science (AAAS) and the American Medical Informatics Association (AMIA).

Matthew P. Page received the M.D. degree from the University of Sydney, Sydney, Australia in 2004 and completed otolaryngology head and neck surgery residency training at the University of Missouri (MU), Columbia, MO, USA in 2010. He is currently an Assistant Professor in the Department of Otolaryngology Head and Neck Surgery, University of Missouri, Columbia, MO, USA. He also heads the MU Voice, Swallow and Airway Center an interdisciplinary training and research program that includes faculty from biomedical, engineering and fine arts departments. His clinical and research interests include voice, swallow and airway disorders.

Guilherme N. DeSouza (S'95–M'01–SM'10) received the Ph.D. degree in electrical and computer engineering from Purdue University, West Lafayette, IN, USA.

He is currently an Associate Professor in the Department of Electrical and Computer Engineering, University of Missouri, Columbia, MO, USA. He also holds adjunct positions with the Department of Computer Science, the MU Informatics Institute and the Sinclair School of Nursing, all at the University of Missouri. He also worked for more than 10 years at the Brazilian Power Systems Research Center on diagnostic of power systems using machine learning and pattern recognition. He has published more than 70 refereed articles in robotic vision, mobile robot navigation, health medicine, and robotic assistive technology. He established the Vision-Guided and Inteligent Robotics Lab in 2005 with funds from the National Science Fundation, Department of Defense, the Leonard Wood Institute, the National Geospatial-Intelligence Agency, the Coulter Foundation, and the MU Research Board. His research is mainly in 3D computer vision and robotic vision.

Dr. DeSouza received the Purdue University's Honeywell Teaching Award, the Maria Canto Neuberger Research Award, and the MU Excellence in Teaching Award. He came to MU after working as a Principal Research Scientist at Purdue University and as a Senior Lecturer at the University of Western Australia.