Dynamical Learning and Tracking of Tremor and Dyskinesia From Wearable Sensors

Bryan T. Cole, *Member, IEEE*, Serge H. Roy, Carlo J. De Luca, *Life Fellow, IEEE*, and S. Hamid Nawab, *Senior Member, IEEE*

***Abstract*—We have developed and evaluated several dynamical machine-learning algorithms that were designed to track the presence and severity of tremor and dyskinesia with 1-s resolution by analyzing signals collected from Parkinson's disease (PD) patients wearing small numbers of hybrid sensors with both 3-D accelerometeric and surface-electromyographic modalities. We tested the algorithms on a 44-h signal database built from hybrid sensors worn by eight PD patients and four healthy subjects who carried out unscripted and unconstrained activities of daily living in an apartment-like environment. Comparison of the performance of our machine-learning algorithms against independent clinical annotations of disorder presence and severity demonstrates that, despite their differing approaches to dynamic pattern classification, dynamic neural networks, dynamic support vector machines, and hidden Markov models were equally effective in keeping error rates of the dynamic tracking well below 10%. A common set of experimentally derived signal features were used to train the algorithm without the need for subject-specific learning. We also found that error rates below 10% are achievable even when our algorithms are tested on data from a sensor location that is different from those used in algorithm training.**

***Index Terms*—Accelerometer signals, dynamical machine learning, electromyographic (EMG) signals, Parkinson's disease (PD), tremor and dyskinesia.**

I. Introduction

PARKINSON'S DISEASE (PD) is one of a number of neuromuscular conditions whose progression can be described in terms of the severity and frequency of associated movement disorders, such as tremor and dyskinesia. Tremor is a periodic involuntary movement, with a frequency of between 3 and 6 Hz, most commonly seen initially in the hands or feet [1], whereas dyskinesia is an aperiodic and uncoordinated type of involuntary movement [2] found in the extremities and trunk. In order to alleviate these conditions, clinicians can titrate anti-Parkinson's medications (such as levodopa) or use surgical implants (for deep-brain stimulation). The efficacy of these treatment methods can be measured by remotely tracking the frequency and severity of the movement disorders. To that end, in current clinical practice patients are typically asked to fill out diaries or questionnaires detailing the type and severity of movement disorders experienced. However, this self-reporting typically does not correlate well with expert annotations in a clinical environment (correlation as low as 0.49 and as high as 0.74 [3]).

Given the proliferation of wearable sensor technology and improvements in machine learning algorithms [4], the potential exists to develop systems that allow clinicians to recognize movement disorders in sensor-wearing patients objectively, remotely, and without intrusion. There have been several notable reports of progress in this direction using a number of different machine learning algorithms.

Salarian *et al.* [5] presented an automatic system based on a Bayesian maximum-likelihood classifier for dynamic tracking of tremor with 1-s resolution using triaxial gyroscope signals from subjects performing scripted sequences of activities such as tooth-brushing and eating. Their algorithm yielded 99.5% sensitivity on tremor-only data and 94.2% specificity on tremor-free data when compared with expert video annotation. This algorithm, however, was not designed to discriminate between tremor and dyskinesia and was therefore not tested on datasets containing instances of dyskinesia.

Keijsers *et al.* [6] used static neural networks (SNN) to track (with a 1-min resolution) dyskinesia by analyzing the signals received from accelerometeric (ACC) sensors worn by PD patients carrying out scripted activities in a randomized order. They reported low error rates when tracking dyskinesia under such conditions. However, in addition to being limited to a 1-min temporal resolution, the tracking method of this report also did not consider the problem of discriminating dyskinesia from tremor.

Patel *et al.* [7] used support vector machines (SVMs) to detect both tremor and dyskinesia throughout the whole body by analyzing the signals from uniaxial ACC sensors worn by PD patients. In order to evaluate algorithm accuracy, patients were asked to perform standardized tests specifically used by clinicians to elicit and assess movement disorders; the SVMs identified whether or not the disorders were present on the basis of features calculated over the entire 30-s duration of each test. Error

Manuscript received June 22, 2013; revised December 16, 2013; accepted February 24, 2014. Date of publication March 19, 2014; date of current version September 04, 2014. This work was supported in part by the National Institute of Health (NIH), under Grant 5 R01 EB007163-05, Grant 5 R01 EB007163-03S1, and Grant 1R43NS083098, as well a grant from the Neuromuscular Research Foundation.

B. T. Cole is with Delsys, Inc., Natick, MA 01760 USA (e-mail: bcole@delsys.com).

S. H. Roy is with the NeuroMuscular Research Center, Boston University, Boston, MA 02215 USA (e-mail: sroy@bu.edu).

C. J. De Luca is with the NeuroMuscular Research Center, the Department of Biomedical Engineering, the Department of Electrical and Computer Engineering, and the Department of Neurology, Boston University, Boston, MA 02215 USA, and also with Delsys, Inc., Natick, MA 01760 USA (e-mail: cjd@bu.edu).

S. H. Nawab is with Department of Electrical and Computer Engineering, Department of Biomedical Engineering, and NeuroMuscular Research Center, Boston University, Boston, MA 02215 USA (e-mail: hamid@bu.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/TNSRE.2014.2310904

1534-4320 © 2014 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

rates of around 2% were reported for the detection of tremor as well as of dyskinesia. It should be noted that this algorithm requires *a priori* knowledge of the specific standardized test that the subject has been asked to carry out during the 30-s interval of ACC signal acquisition. Furthermore, this algorithm also has the practical limitation that before it can be used to detect the presence of tremor or dyskinesia in a subject, it must first be trained on previous data acquired from the *same* subject.

Palmes *et al.* [8] utilized an ensemble of several cooperating SVMs to analyze surface electromyographic (EMG) sensor data to recognize the presence of tremor in PD patients. Their algorithm was able to distinguish scripted activities performed by PD patients from those performed by healthy controls with an error rate of 2% using subject-independent training. Besides producing only one decision per scripted activity, their algorithm's applicability is limited only to the detection of tremor.

Motivated by the encouraging results from previous work by others, we set out to overcome the limitations of these other algorithms. Most importantly, our goal was to develop algorithms that can track *both* tremor and dyskinesia *without restricting* the patient to scripted or standardized activities. Rather, we envisioned the patient freely carrying out activities of daily life in an apartment/home environment while wearing wireless body-worn sensors. The algorithms would monitor the resulting sensor signals for the purpose of establishing the presence/absence of tremor and dyskinesia and assessing their respective severity levels in accordance with generally accepted clinical criteria.

In this paper, we report our success in developing such algorithms by comparing the tracking capability of three different dynamical machine-learning frameworks: dynamic neural networks (DNNs), dynamic support vector machines (DSVMs), and hidden Markov models (HMMs). We show that the choice of dynamic rather than static machine learning algorithms allows us to exploit the time-varying nature of tremor and dyskinesia to accurately detect these disorders in the presence of unscripted and unconstrained voluntary movements.

II. Algorithm Design Specifications

In this section, we summarize the functional specifications and signal processing tasks for the algorithm.

A. Functional Specifications

Sensor Data: The algorithm will process multiple sources of data acquired concurrently from wearable sensors. The sensor data will include both information about muscle activation via surface electromyographic (EMG) sensors, as well as body movement and position as tracked by triaxial accelerometers (ACC). The algorithm will also require a minimal number of sensors on the body and be relatively insensitive to minor differences in sensor placement that result from anatomical differences between subjects or test/re-test applications.

Disorder Tracking: The algorithm will provide a temporal history of the presence of tremor and dyskinesia, as well as a severity rating, based on standardized clinical outcome measures derived through expert annotation of video recordings. The beginning and end of each movement disorder occurrence will be identified with a resolution of 1 s to enable clinicians to closely monitor rapid changes in movement disorder severity, which are a common complication of advanced PD.

Activity Conditions: The algorithm will process sensor data acquired during unconstrained activities of daily living in the home environment. It will function equally well during sitting, standing and walking mobility states. This requirement avoids the inconvenience of imposing scripted tasks or standardized test movements on the subject, and thus best approximates the intended clinical usage of the algorithm. A preponderance of the movement disorder studies that have resulted in wearable monitoring solutions have been constrained to these operating conditions so as to simplify the tracking task of the algorithm [5]–[8].

Training: The algorithm will track movement disorders in subjects whose sensor data did not contribute to the training phase of the algorithm. This specification eliminates the need for extensive and costly retraining for each new subject. While some previous algorithms for tracking individual disorders (e.g., for tracking tremor in [5]) have possessed this attribute, more recent algorithms for tracking multiple disorders (e.g., for tracking tremor and dyskinesia in [7]) required additional training for each new subject.

Error Rate: The specification for an upper limit of allowable error rates is set at 10%. Error rates will be derived based on the widely used procedure of comparing algorithm outputs to expert annotations of video recordings [5], [7]. Note that we tolerate a wider range of error rates than that achieved in previously reported studies (e.g., in [7] and [8]) because our algorithms will operate under significantly less constrained conditions than these others.

B. Signal Processing Tasks

The signal processing solutions will need to resolve the following three tasks.

Distinguishing Tremor From Voluntary Movements: Accelerometer and EMG recordings of tremor are manifested by periodic activity in the 4–6 Hz frequency range. Although most typical voluntary movements do not present with such characteristic frequencies, evidence of tremor can be attenuated or obscured in the presence of voluntary movements, as demonstrated for the signal from the ACC sensor in Fig. 1. In these situations, the signal processing routines must overcome the task of detecting the underlying periodicity of tremor despite the contaminating influence of voluntary movement in one or both of the EMG and ACC sensor modalities.

Distinguishing Dyskinesia From Voluntary Movements: A key signal processing task is to distinguish between rapid changes in the ACC recordings caused by dyskinesia from those produced by rapid voluntary movements. Fig. 2 provides an example of one such type of movement. Here, ACC sensor signals recorded from a patient feeding himself in the absence of dyskinesia are compared to ACC sensor signals from the same limb during dyskinesia. Fig. 2(a) demonstrates that both types of movements produce similar waveforms. However, closer inspection of the corresponding power spectral densities [Fig. 2(b)] demonstrates that dyskinetic movements may have significant frequency components as high as 8 Hz [2], whereas

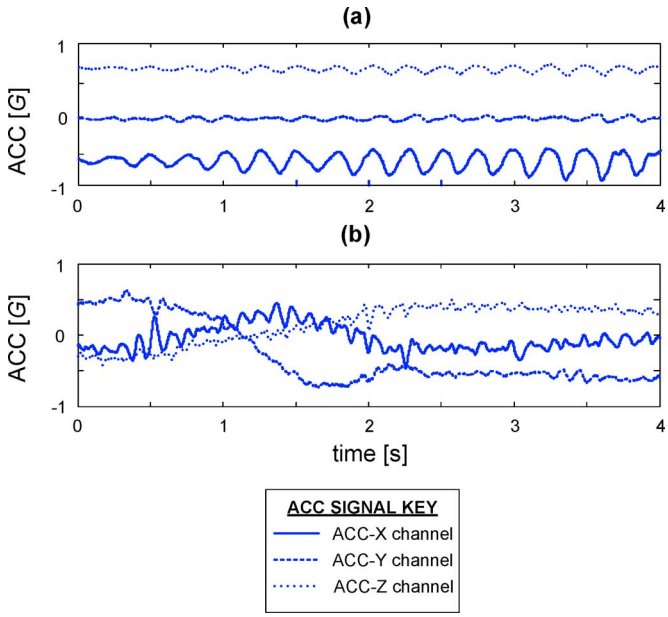


Fig. 1. (a) Triaxial ACC signals recorded from the wrist extensor site during *resting* tremor while sitting, and (b) ACC sensor signals collected from the wrist extensor site during *kinetic* tremor recorded while the patient was voluntarily moving his arm. In the absence of movement as in (a), the identification of tremor is straightforward and depends on the recognition of the periodicity that characterizes movements due to tremor; however, the presence of the much slower and more energetic voluntary movement in (b) can obscure the periodicity of tremor.

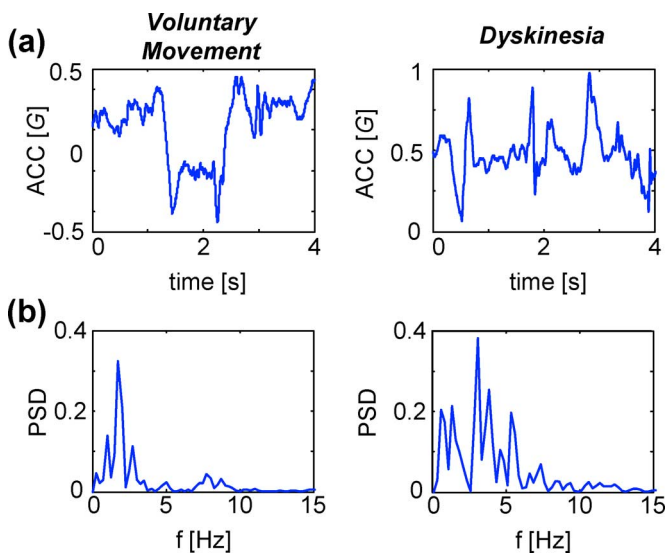


Fig. 2. (a) One channel of the ACC signals recorded by the sensor above the wrist extensor muscle while the subject performed voluntary movement and experienced dyskinesia, respectively. (b) The power spectral densities associated with the signals shown in (a). From these PSDs, we can tell that energy during dyskinesia is concentrated in the higher frequencies, whereas energy during voluntary movement is concentrated in the lower frequencies.

voluntary movements do not. This key difference will be exploited by the signal processing routines when developing the tracking algorithm for dyskinesia.

Distinguishing Tremor From Dyskinesia: A key signal processing task is to distinguish between the near-sinusoidal 3–6 Hz characteristic of the tremor ACC signal from the generally aperiodic and higher frequency content of the dyskinesia

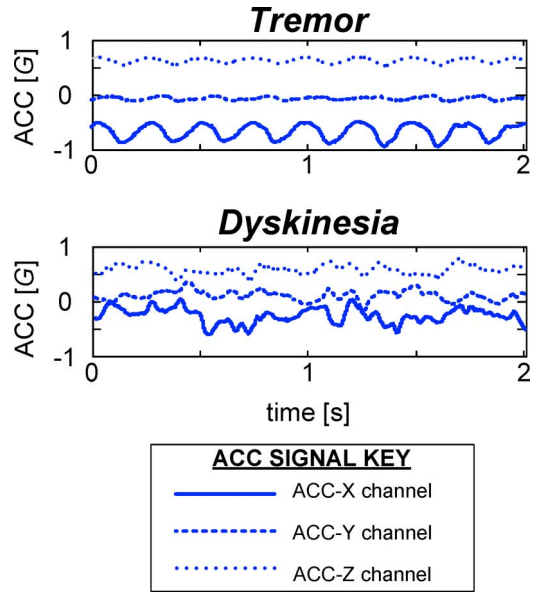


Fig. 3. Signals recorded from the ACC sensors above the wrist extensor muscle in patients experiencing tremor (top row) and dyskinesia (bottom row). In the ACC sensor signals, tremor presents as a sinusoidal wave with periodicity in the 3–6 Hz range, whereas dyskinesia produces aperiodic movements with no fundamental frequency.

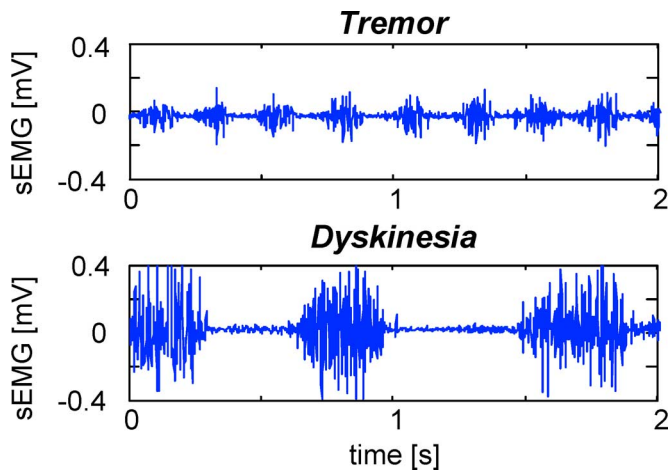


Fig. 4. Signals recorded from the ACC and EMG sensors above the wrist extensor muscle in patients experiencing tremor (top row) and dyskinesia (bottom row). In the sEMG sensor signals, tremor produces regular bursts of energy at the same fundamental frequency seen in the ACC signals, whereas dyskinesia is represented by sporadic bursts of energy that vary in amplitude, duration, and time between bursts.

ACC signal (see Fig. 3). For the EMG sensor data, the task is similar: the periodic bursts associated with tremor must be distinguished from the more randomly distributed noise-like bursts of dyskinesia (see Fig. 4).

III. Dynamic Learning Structures

Our approach to tracking tremor and dyskinesia is twofold: first, design a *dynamic* learning algorithm that can track the presence and absence of these disorders using EMG and ACC sensor data; and second, track their severity by analyzing filtered versions of the ACC signal energy. We implemented dynamic learning algorithms rather than their *static* counterparts in order

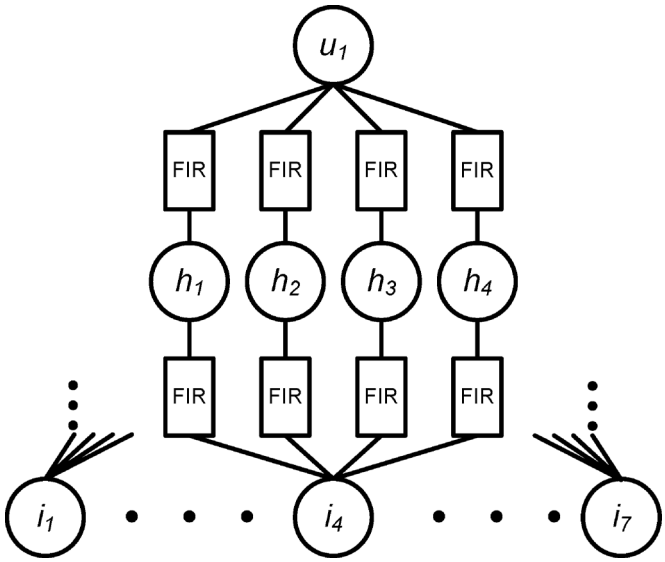


Fig. 5. Diagram of the DNN used to recognize tremor. There are seven input nodes, four hidden nodes, and one output node. Each FIR filter in the hidden layer has five weights, whereas each filter in the output layer has one weight. The DNN used to recognize dyskinesia is structured similarly, with five input nodes, two hidden nodes, and one output node.

to enable the algorithm to learn not only patterns of signal variation associated with the involuntary movements, but also to learn how those patterns may *change over time*. The superposition of voluntary and involuntary movements necessitates such an approach: a disorder clearly present at one point in time may be obscured by voluntary movement in the next.

We adapted three types of dynamic learning structures: DNNs, DSVMs, and HMMs. Each of the three types represents a different approach to the classification problem: HMMs classify sequences of features probabilistically, whereas DNNs divide the feature space using a series of linear hyperplanes, and DSVMs a series of nonlinear hyperplanes.

For each approach, we developed two separate classifiers—one to detect tremor and one to detect dyskinesia—to better address the unique challenges presented by the identification of each disorder. Below, we discuss how we adapted each of these structures to track tremor and dyskinesia. In later sections we compare their performance to determine whether any one structure is superior for these applications.

A. Dynamic Neural Networks (DNNs)

Two DNNs were used to detect the presence/absence of tremor and dyskinesia, respectively. Both networks were trained using backpropagation with learning rate $\alpha = 0.05$ over 1000 iterations.

1) DNNs for Tremor: We developed a two-layered dynamic neural network (DNN) with one hidden layer and an output layer for tracking the presence/absence of tremor, as illustrated in Fig. 5. Each of the four hidden nodes, h_i in Fig. 5, applies the weights of a five-point FIR filter to time-advanced and time-delayed versions of four of the ACC-based and all three of the EMG-based signal features discussed in Section IV.

The values of the coefficients associated with each FIR filter were learned during the training phase. The training epochs were carefully chosen to be representative of the different manifestations of both tremor and dyskinesia during unscripted activity. Each epoch consisted of 6 s of signal data collected from the hybrid sensor placed above a particular muscle on the dominant limb. For the DNN designed to detect tremor in the arm, the sensor was placed above the wrist extensor muscle; for the DNN designed to detect tremor in the leg, the sensor was placed above the tibialis anterior muscle in the shin. The choice of these particular sensor sites was based on the fact that, in our database as a whole, tremor and dyskinesia tended to manifest most often at these particular sites; we also found the dominant limb tends to exhibit a larger variety of voluntary movements.

2) DNN for Dyskinesia: We developed a two-layered network with two hidden nodes to track dyskinesia, similar to the DNN designed to recognize tremor shown in Fig. 5. Each hidden node applies the weights of a five-point FIR filter to time-advanced and time-delayed features calculated from the input data. Only the five ACC-based features (see Section IV) were used as inputs because we found it difficult to reliably discern differences between EMG signals corresponding to dyskinesia and those corresponding to voluntary movements. The dyskinesia DNN was trained with the same training data as the tremor DNN.

B. Dynamic Support Vector Machines

As was the case for DNNs, two separate DSVMs were designed to detect tremor and dyskinesia, respectively.

1) DSVM for Tremor: The tremor DSVM was designed to use the same input signal features as our tremor DNN and to also have the same level of computational complexity. We also used the same training data for the tremor DSVM as for the tremor DNN. We investigated the use of a variety of different kernel functions, specifically linear, polynomial, sigmoid, and radial basis functions. For each kernel function, we adjusted the associated parameters, and determined the global error rate over the training dataset. Because there are many choices of parameters and kernel functions that will produce very low error rates over the training database, additional cross-validation was required to determine which choice of parameters and kernel function will produce the DSVM transformation that best generalizes over the testing dataset. As a result of these processes, we selected a DSVM transformation that incorporated a radial basis function with a trade-off coefficient C of 1 and scale factor γ of 0.25; the final DSVM transformation included 86 support vectors.

2) DSVM for Dyskinesia: The dyskinesia DSVM was designed to use the same input signal features as our dyskinesia DNN and to also have the same level of computational complexity. We also used the same training data for the dyskinesia DSVM as for the dyskinesia DNN. Our results led us to select a DSVM transformation that used a sigmoid basis function with a trade-off coefficient C of 0.125 and scale factor γ of 0.5; the final DSVM transformation included 80 support vectors.

C. Hidden Markov Models

Whereas DSVMs and DNNs seek to find the hyperplane that best divides the feature space associated with two signal classes, HMMs seek to find the probability distribution that governs the

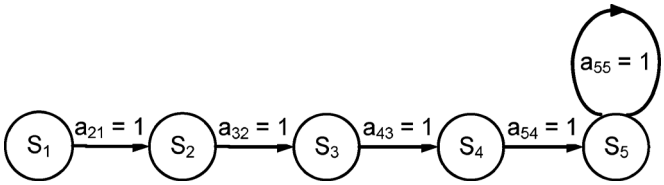


Fig. 6. Diagram of the HMMs used to recognize tremor and nontremor segments in the isolated word recognizer framework from Rabiner [12]. Model is a left-right model with five total states. Number of states was fixed at the number of windows, such that moving from one window to the next always represented a state transition.

input patterns produced by each signal class. HMMs thus represent a more statistical approach to classification.

We used the HMM-based isolated word recognizer framework described by Rabiner [12] as the starting point for addressing the dyskinesia and tremor tracking problems. In this framework, two competing HMMs were created: one to describe intervals containing the phenomenon of interest (e.g., tremor), and the other to describe intervals not containing this event. For each new interval with unknown disorder state, signal features are calculated and each model produces a probability of observing those features conditional upon either the absence or presence of the disorder. Depending upon which model produces the higher probability, the phenomenon of interest is declared to be present or absent.

In our case, each disorder-present and disorder-absent HMM utilizes features computed over five consecutive 2-s windows of the data, each with 1-s overlap with the next one. A hidden state (presence or absence of disorder) is represented by the signal at each window position and moving from one window position to the next can be thought of as representing a state transition, as illustrated in Fig. 6. Given such a model λ for the presence of tremor, suppose we wanted to know the probability that the features computed from five consecutive window positions (with middle-window position at time k) represent the observation of a tremor process. Under the Markovian assumption, we can formulate this probability as a product of five observation probabilities as derived from a training data set.

1) HMMs for Tracking Tremor: The observation probabilities [12] associated with both the tremor and nontremor models were computed using the training data. The features were individually quantized into two bins, with the cutoff for each bin set by a machine learning classifier dividing the range of possible values between those that were more likely to indicate tremor and those more likely to indicate nontremor. This procedure was performed for each of the features used by our tremor HMM-based algorithm to divide the seven-dimensional feature space into $2^7 = 128$ possible symbols.

2) HMMs for Tracking Dyskinesia: As with tremor, we created two HMMs: one to describe intervals containing dyskinesia, and the other to describe intervals not containing dyskinesia. Features for any particular observation time were computed using the same windowing strategy as for the tremor HMMs. Feature quantization was also done in the same way as for the tremor models except that there were only five features, all derived from the ACC signals. In total, the five-dimensional feature space was divided into $2^5 = 32$ possible symbols.

IV. Calculation of Baseline Signal Features

In total, eight signal features were computed from the ACC and EMG sensor signals and used as inputs to the dynamic machine learning algorithms designed to identify the absence/presence of tremor and dyskinesia. Note that the tremor detection algorithms used one subset of features, and the dyskinesia detection algorithms used a second subset.

All features were calculated over a 2-s window, with 1-s overlap between adjacent windows. The features are described in detail below.

A. ACC Lowpass Energy (A_{LP})

A_{LP} is the lowpass energy *below* 1 Hz in an ACC epoch after subtraction of the mean epoch value to remove the gravity component. A high value (in accordance with a learned threshold) for A_{LP} may be used as an indicator of the possible presence of significant voluntary activity since neither tremor nor dyskinesia contributes significantly to this portion of the ACC spectrum. We experimented with a variety of cutoff frequencies (ranging from 0.5 to 3 Hz in 0.25 Hz increments) for this feature but found that our learning algorithms gave their best performance with a 1 Hz cutoff. This feature is used in our algorithms for *both* tremor and dyskinesia.

B. ACC Highpass Energy (A_{HP})

A_{HP} is the energy *above* 1 Hz in an ACC epoch. A high value (in accordance with a learned threshold) for A_{HP} may be used as an indicator of the possible presence of either tremor or dyskinesia. Again, we experimented with a variety of cutoff frequencies (ranging from 0.5 to 3.0 Hz in 0.25 Hz) but found that our learning algorithms gave their best performance with a 1 Hz cutoff. This feature is used in our algorithms for *both* tremor and dyskinesia.

C. ACC Very-Highpass Energy (A_{VHP})

A_{VHP} is the energy *above* 15 Hz in an ACC epoch. A high value (in accordance with a learned threshold) for A_{VHP} may be used as an indicator of the possible presence of dyskinesia since neither tremor nor most voluntary movements are anticipated to contribute significantly to this feature. We experimented with a variety of cutoff frequencies (ranging from 10 to 20 Hz in 0.5 Hz increments) for this feature but found that our learning algorithms gave their best performance over the training data set with a 15 Hz cutoff. This feature is used *only* in our algorithms for dyskinesia.

D. ACC Lag of First Autocorrelation Peak (A_L)

A_L is defined as the location of the first positive side lobe of the autocorrelation of an ACC epoch provided there are also additional peaks in the autocorrelation at one or more positive integer multiples of that lag; otherwise A_L is set to zero. The location of the first positive side lobe in the autocorrelation serves as an indicator of the fundamental frequency of the underlying signal and was used to detect the presence of periodicity in an epoch. This feature is used in our algorithms for *both* tremor and dyskinesia.

E. ACC Height of First Peak in Autocorrelation (A_H)

A_H is defined as the ratio of the height of the first positive side lobe in the autocorrelation of an ACC epoch to the height of the main lobe found at zero lag. A high value for A_H may be used as an indication that at least two significant portions of the epoch are similar in shape to each other. Along with A_L, this feature can be used to help discriminate between the periodic characteristics of tremor from the aperiodic characteristics of both dyskinesia and many voluntary movements. This feature is used in our algorithms for *both* tremor and dyskinesia.

F. EMG Energy (E_E)

E_E is a measure of the root mean square energy in an EMG epoch. Since significant energy in the EMG signal is an indicator of force generation in a specific muscle [13], a high value for E_E may be used to suggest that movement may be taking place. However, it is important to note that if a movement occurs without the force contribution from a particular muscle, an EMG sensor placed over that muscle will not detect evidence of movement. This feature is used *only* in our algorithms for tremor.

G. EMG Lag of First Peak in Autocorrelation (E_L)

E_L is defined as the location of the first positive side lobe of the autocorrelation of an EMG epoch and was computed in a manner similar to that of A_L described above. This feature is used *only* in our algorithms for tremor because the EMG signal produces repeated bursts of energy approximately every 1/3 to 1/6 of a second, which are identifiable by the strong positive peaks in the autocorrelation function. In contrast, dyskinetic movements do not produce such periodic bursts of energy.

H. EMG Height of First Peak in Autocorrelation (E_H)

E_H is defined as the ratio of the height of the first positive side lobe in the autocorrelation of an EMG epoch to the height of the main lobe (at zero lag) in the autocorrelation. The empirically verified assumption behind this feature is that a higher proportion of the EMG signal's energy tends to be concentrated at the first side lobe of the autocorrelation for periodically repeating bursts of energy as seen in epochs containing tremor. This is in contrast to the EMG signal energy observed in the presence of aperiodic movements, including dyskinesia and most voluntary movements. This feature is used *only* in our algorithms for tremor.

V. Algorithm Performance Metrics

Discrepancies between the output of the classification algorithm and expert annotation were evaluated on the basis of sensitivity (which describes the ability of the algorithm to correctly identify a movement disorder when it is present) and specificity (which describes the ability of the algorithm to correctly identify all instances when the movement disorder is absent) [14]. Other studies [8] also use accuracy as a single metric to compare the performance of their algorithms. Defined as the as the ratio of the number of correct decisions (true positives and true negatives) to the total number of decisions, accuracy can alternately be written as

$$\text{accuracy} = (F_p \times Sn) + (F_a \times Sp).$$

Here, F_p is the fraction of the total seconds in the database in which the disorder is present, Sn is algorithm sensitivity, F_a is the fraction of total seconds in the database in which the disorder is absent, and Sp is algorithm specificity.

From this formulation we can observe the dependence of the accuracy on the proportion of disorder present. This may be acceptable to judge a classifier over a database containing equal amounts of all disorders of interest. However, for our database of unscripted and unconstrained activities, we found unequal proportions of tremor, dyskinesia, and intervals with no disorder (as we will see in Section VI). We also found that more sedentary activities (such as reading, talking, or using a computer) were more common than more vigorous activities (such as walking). Reliance on accuracy could thus lead to biased judgments of performance that favored algorithms which overreported the more common disorders, or which underperformed during certain movement states.

With this consideration in mind we therefore developed a *global* error rate (GER) in detecting a particular disorder within the entirety of a given signal database. Global error rate is derived from the ratio of the number of incorrect decisions (false positives and false negatives) to the total number of seconds of data in the signal database, i.e., $\text{GER} = 1 - \text{accuracy}$. However, to avoid the possible biases discussed previously, we assume that the sensitivity and specificity values have stabilized over the given database in such a way that if the amount of the data in the database were increased, these values would remain the same. Under this assumption, we can now use $F_p = F_a = 0.5$ for a conceptual extension of the original database and arrive at the following general definition for the global error rate:

$$\text{GER} = 1 - (0.5 \times Sn) - (0.5 \times Sp).$$

However, further normalizations to GER are required to account for the number of potential disorder states. Consider the case of calculating Sp for a tremor-tracking algorithm. This calculation would be made over intervals with no disorder whatsoever as well as over intervals with dyskinesia. If the likelihoods of error by the algorithm in these two types of intervals are significantly different, the overall sensitivity of the algorithm will change as the relative frequency (in the database) of intervals without disorders is changed with respect to intervals with dyskinesia. Thus, the global error rate for this tremor tracking algorithm should be calculated

$$\text{GER}_{\text{norm}} = 1 - (0.5 \times Sn) - (0.25 \times (Sp_N + Sp_D)).$$

Here, Sp_N is the specificity calculated over database intervals where no disorder is present and Sp_D is the specificity over intervals where the other disorder is present. We have now normalized GER with respect to relative disorder frequency.

Lastly, we find that both sensitivity and specificity values for tremor and dyskinesia are highly dependent on the subject's

broad *mobility state* classification, i.e., on whether the subject is walking, standing, sitting, or lying down. For example, the sensitivity of a tremor-tracking algorithm can be significantly higher when the subject is sitting down in comparison to when the subject is ambulating. As previously stated, the relative frequency of these movement states in our database is not necessarily representative of the relative frequency of movement states in other PD patients. By normalizing the relative frequency of these movement states, we produce the final definition of the GER used to describe algorithm performance

$$\mathrm{GER} = 0.25(\mathrm{GER}_{\mathrm{walk}} + \mathrm{GER}_{\mathrm{stand}} + \mathrm{GER}_{\mathrm{sit}} + \mathrm{GER}_{\mathrm{lie}}).$$

Here, $\mathrm{GER}_{\mathrm{walk}}, \mathrm{GER}_{\mathrm{stand}}, \mathrm{GER}_{\mathrm{sit}}$, and $\mathrm{GER}_{\mathrm{lie}}$, respectively represent the calculation of $\mathrm{GER}_{\mathrm{norm}}$ over only the data where the subject is walking, standing, sitting or lying down.

Achieving an acceptably low global error rate does not always mean that all decisions made by the corresponding disorder tracking algorithm are credible. For example, poor algorithm performance in the presence of a specific mobility state can manifest in dense groupings of errors. Therefore, we also devised an additional error rate—which we call the local error rate (LER)—to measure the performance of our algorithms in errors that are "too dense." For the purposes of this paper, we define the LER as the percentage of all possible 30-s intervals defined over the database that have more than 15 errors within them. For example, an algorithm would have an LER of 5% when 5% of all 30-s intervals defined on a database have an error rate of at least 50% and thus contain what we consider to be dense errors.

VI. Data Acquisition

A. Subjects

Two groups of subjects were tested (Table I): one ($\mathrm{n} = 11$ with PD) provided a training data set for algorithm development and the other ($\mathrm{n} = 8$ with PD; $\mathrm{n} = 4$ controls) provided a data set for testing the algorithms. The acquisition of separate databases was implemented to demonstrate that the algorithms are subject-independent and need not require pre-training for each application. All of the PD patients were screened for mild to moderately severe categories of Parkinson's disease (Hoehn-Yahr stages II–III while "on" and Hoehn-Yahr stages III–IV while "off") [15], complicated by motor fluctuations that included mild to severe ranges of tremor and dyskinesia severity. The control subjects were selected to be within the age range of the patients and were screened for neuromuscular disorders, including PD. Voluntary written informed consent was obtained from each participant, in compliance with institutional review board procedures.

B. Sensor and Sensor Placement Specification

Eight hybrid sensors (DelSys, Inc.) were placed at various locations on the subjects' arms, legs, and sternum as illustrated in Fig. 7. Each of these hybrid sensor acquires and wirelessly transmits three channels of triaxial ACC data and one channel of EMG data, all sampled at 1 kHz (with appropriate anti-aliasing filtering). A photograph of one of the sensors is provided in the insert in Fig. 7. These sensors were designed for practicality and ease of patient use while performing everyday activities. In a real-world clinical application, however, it is desirable to have each subject wear as few sensors as possible.

TABLE I
Subject Population Characteristics

	Training Set	Testing Set
Controls	n = 0	n = 4
Age (yr)		54±16.6
Males/Females		4 M/0 F
Patients	n = 11	n = 8
Age (yr)	61.1±5.5	62.9±5.3
Males/Females	9 M/2 F	7 M/1 F
Disease duration (yr)	13.5±6.0	13.2±9.2
Tremor prevalence (%)	17.7±19.3	16.7±20.7
Tremor duration (s)	39.4±43.0	42.3±34.4
Dyskinesia prevalence (%)	49.7±45.1	48.5±25.7
Dyskinesia duration (s)	52.3±51.7	62.6±45.6

Characteristics of the subject populations used for training and testing the dynamic machine learning algorithms. "Prevalence" denotes the percentage of the recording period with the disorder present, regardless of severity or body location. "Duration" denotes how long a disorder persisted at a particular severity level.

Initial feature extraction and algorithm development [14] led us to conclude that it was possible to adequately track tremor and dyskinesia using data from just *two* sensor locations. Those sensors [denoted by stars in Fig. 7] are placed over the wrist extensor muscle of the dominant symptomatic arm and the tibialis anterior muscle in the shin of the dominant symptomatic leg. Results reported in this study are derived from these two locations.

C. Protocol

All of the subjects carried out unscripted and unconstrained activities of daily living in a simulated home environment for a continuous period of 3–4 h per subject to capture a complete "On–Off" medication cycle for each patient. Experimental sessions were continuously videotaped using fixed and hand-held high resolution digital cameras, and then stored for later annotation. Video and sensor data were synchronized by generating a pulse tone recorded on the cameras audio channel. Sessions were timed to begin approximately 1 h following the patient's first morning dose of anti-Parkinsonian medication. The subjects were free to move about the environment without coaching from the researchers or use of an activity script. Activities were varied and included periods of sitting, standing, walking, resting (lying down on a bed), preparing snacks, eating, reading, writing, using the Internet, and conversing with researchers and family members.

D. Annotation

Video annotation for scoring tremor and dyskinesia severity was carried out by a team of movement disorder specialists consisting of two neurologists, a nurse coordinator (both from the Boston University Parkinson's Disease Center), and a physical therapist. This information provided the basis by which the classification algorithms were trained and tested. Tremor severity was scored based on Item 20 (Tremor at Rest) and Item 21 (Action or Postural Tremor) of the Motor Examination section of the

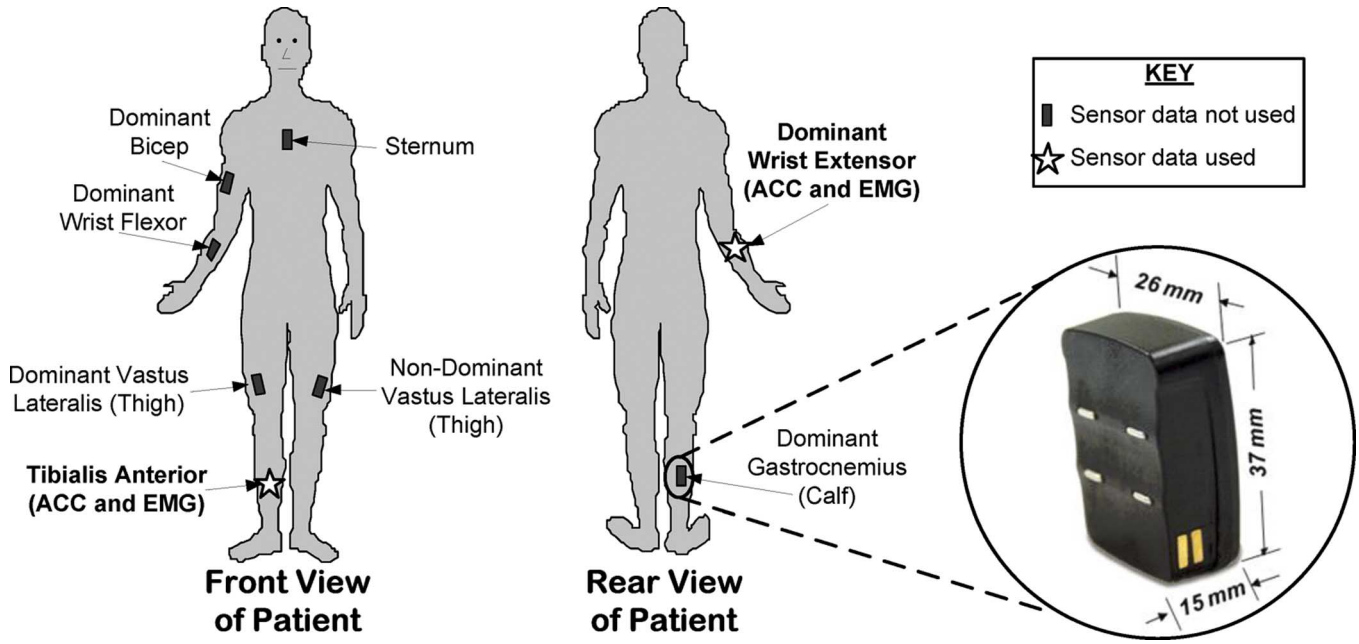


Fig. 7. Locations of the eight wireless sensors worn by the subjects in our studies. The classifiers described in this paper rely on data from one hybrid sensor (denoted by stars) to track the evolution of both tremor and dyskinesia in a particular limb. A picture of one of these wireless sensors is visible in the inset. The sensor collects three channels of data from a triaxial accelerometer and one channel of surface electromyographic data.

Unified Parkinson's Disease Rating Scale (UPDRS) [15]. Dyskinesia severity was scored based on the m-AIMS scale [16]. Both instruments use a five-point Likert scale where 0 corresponds to the absence of the disorder and 4 correspond to the most extreme disorder. Annotators identified the beginning and end of each movement disorder severity occurrence with a resolution of 1 s, re-examining those sections where their annotations disagreed. If no agreement was reached, the data point was discarded. Each of the four limbs was scored separately.

VII. Results

In this section, we discuss the performance of our algorithms for tracking the presence and severity of both tremor and dyskinesia.

A. Tracking Presence and Severity of Tremor

Table II summarizes the performance of our tremor-tracking algorithms for three of the basic mobility states, based on DNN, DSVM, and HMM frameworks. The performance of the algorithms was evaluated in terms of the global and local error rates that were described in Section V. We note that all the results listed in Table II pertain to the case where all the testing data was from the hybrid ACC-EMG sensor placed above the wrist extensor muscle of the dominant arm; this is the same sensor location used to train the various machine-learning algorithms.

Table II provides a summary of the comparative results of the machine learning algorithms demonstrating that they are approximately equally effective at detecting tremor in the presence of a variety of basic mobility states, with a slight advantage in favor of the DNN algorithm when both GER and LER are considered.

Because the severity of movement disorders is important to the clinician, we have also developed pattern classification techniques to identify all seconds in which tremor is recognized as containing mild, moderate, or severe tremor. The severity level estimation is based on a simple Bayesian maximum likelihood classifier [4] applied to the ACC highpass energy (A_{HP}) feature. Our severity classifier used the category of "mild" tremor for UPRDS rating of 1; "moderate" tremor for UPRDS rating of 2; "severe" tremor for UPRDS severity ratings of 3 *and* 4.

TABLE II
Comparison of Tremor Recognition in Dominant Arm by DNN, DSVM, and HMM Algorithms

Movement State	DNN		DSVM		HMM	
	GER	LER	GER	LER	GER	LER
Sitting	4.8%	0.23%	4.2%	0.75%	4.0%	0.46%
Standing	8.9%	0.22%	13.2%	0.11%	11.0%	0.16%
Walking	4.9%	0.37%	4.1%	0.37%	3.2%	0.74%
All States	6.2%	0.28%	7.2%	0.41%	6.1%	0.45%

Comparison of performance of the DNN, DSVM, and HMM algorithms for tracking tremor over the 44-hour testing database. Performance is measured in terms of global error rate (GER) and local error rate (LER).

Table III shows that this Bayesian maximum likelihood classifier can be used to accurately distinguish between the different severity levels of tremor by achieving a sensitivity and a specificity of greater than 95% for each severity level.

B. Tracking Presence and Severity of Dyskinesia

Table IV summarizes the performance of our machine learning algorithms designed to track dyskinesia. The results have the same general trends as those in Table II; we see again that all three classifiers are of approximately equal efficacy.

As with tremor, we applied a simple Bayesian maximum likelihood classifier to A_{HP} to measure the severity of any detected instances of dyskinesia. Table V shows that this classifier can be used to distinguish between the different severity levels of

TABLE III
TREMOR SEVERITY RECOGNITION PERFORMANCE OF BAYESIAN MAXIMUM LIKELIHOOD CLASSIFIER

Category	Sens	Spec	Hours of Data	Number of Subjects
Mild Trem	97.2%	97.8%	1:56	5
Mod Trem	95.2%	97.1%	1:14	4
Sev Trem	96.3%	99.3%	0:26	3

Performance of the Bayesian maximum likelihood classifier for tremor severity over the 44-hour testing database. The sensitivity (Sens) and specificity (Spec) were calculated for each of the three levels of tremor: mild, moderate (Mod), and severe (Sev).

TABLE IV
COMPARISON OF DYSKINESIA RECOGNITION IN DOMINANT ARM BY DNN, DSVM, AND HMM ALGORITHMS

Movement State	DNN		DSVM		HMM	
	GER	LER	GER	LER	GER	LER
Sitting	6.4%	1.6%	6.5%	4.7%	12.2%	2.4%
Standing	6.8%	3.9%	10.2%	7.0%	13.1%	3.7%
Walking	13.1%	9.0%	10.5%	7.4%	11.5%	6.3%
All States	8.8%	4.8%	9.1%	6.4%	12.3%	4.2%

Comparison of performance of the DNN, DSVM, and HMM algorithms for tracking dyskinesia over the 44-hour testing database. Performance is measured in terms of global error rate (GER) and local error rate (LER).

TABLE V
DYSKINESIA SEVERITY RECOGNITION PERFORMANCE OF BAYESIAN MAXIMUM LIKELIHOOD CLASSIFIER

Category	Sens	Spec	Hours of Data	Number of Subjects
Mild Dys	93.9%	95.5%	0:52	5
Mod Dys	91.9%	94.6%	0:46	5
Sev Dys	95.0%	98.6%	0:13	4

Performance of the Bayesian maximum likelihood classifier for dyskinesia severity over the 44-hour testing database. The sensitivity (Sens) and specificity (Spec) were calculated for each of the three levels of tremor: mild, moderate (Mod), and severe (Sev).

dyskinesia (mild, moderate, and severe) with a sensitivity and specificity of greater than 92% for each severity level.

C. Tracking at Other Sensor Locations

Because symptoms of PD can present differently in different limbs at the same time, we conducted a further analysis to determine how our dynamic learning algorithms, which were trained using sensor data acquired from the location corresponding to the wrist extensor muscle, would perform on sensor data from a different location, corresponding to the tibialis anterior (TA) muscle. Specifically, we tested the DNN algorithms using the TA sensor data from the entire testing database. In the case of tremor tracking, we found a GER of 8.4% and LER of 3.6%. In the case of dyskinesia tracking, we found a GER 10.7% and a LER of 2.2%

VIII. DISCUSSION AND CONCLUSION

We have demonstrated that the machine learning algorithms developed in this study were approximately equally effective in tracking tremor and dyskinesia with a 1-s temporal resolution during unscripted and unconstrained activities in a home-like environment. Data were derived from a relatively small number of hybrid ACC-EMG sensors worn by PD patients and controls carrying out routine activities of daily living. In particular, we found that global and local error rate performance metrics for DNN, DSVM, and HMM algorithms averaged below 10%, with the DNN algorithm achieving the best overall metrics. None of these algorithms needed to be retrained using additional data from the test subjects. The tremor and dyskinesia algorithms are computationally simple enough so that their net execution time on a typical commercially available laptop is of the same order of magnitude as the duration of their input signals, thus demonstrating that real-time implementations of the algorithms are feasible.

These results support the likelihood of developing an ambulatory system for monitoring fluctuations in tremor and dyskinesia in patients with PD. This achievement will provide significantly greater temporal resolution and objectivity than is possible through the current standard of self-report measures such as motor diaries. The proliferation of unobtrusive yet objective measures of Parkinsonian symptoms should enable health care providers to more effectively titrate anti-PD treatments, thus improving the treatment of such movement disorders without unduly restricting the daily lives of their patients.

ACKNOWLEDGMENT

The authors are grateful to all participants for their time and patience during testing. The authors would like to thank M. M. Saint-Hilaire MD, FRCP, C. A. Thomas, and other clinical staff at the Boston University (BU) Parkinson's Disease & Movement Disorders Center for assisting with patient referrals, scheduling, and video annotation. The authors would also like to thank the following BU students for their time and effort in processing the large data sets produced by the study: N. Elkayam, H. Younis, M. Searle, M. Wexler, S. Ganesan, P. Ozdemir, F. Shah, and L. Matabusi.

REFERENCES

[1] R. J. Elble, "Tremor: Clinical features, pathophysiology, and treatment," *Neurol. Clin.*, vol. 27, no. 3, pp. 679–695, 2009.
[2] J. Jankovic, "Motor fluctuations and dyskinesias in Parkinson's disease: Clinical manifestations," *Mov. Disord.*, vol. 20, no. 11, pp. S11–S16, 2005.
[3] J. Reimer, M. Grabowski, O. Lindvall, and P. Hagell, "Use and interpretation of on/off diaries in Parkinson's disease," *J. Neurol., Neurosurg., Psychiatry*, vol. 75, pp. 396–400, 2004.
[4] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley-Interscience, 2001.
[5] A. Salarian *et al.*, "Quantification of tremor and bradykinesia in Parkinson's disease using a novel ambulatory monitoring system," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 2, pp. 313–322, Feb. 2007.
[6] N. L. Keijsers, M. W. Horstink, and S. C. Gielen, "Automatic assessment of levodopa-induced dyskinesias in daily life by neural networks," *Mov. Disord.*, vol. 18, no. 1, pp. 70–80, Jan. 2003.
[7] S. Patel *et al.*, "Monitoring motor fluctuations in patients with Parkinson's disease using wearable sensors," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 6, pp. 864–873, Nov. 2009.
[8] P. Palmes, W. T. Ang, F. Widjaja, L. C. S. Tan, and W. L. Au, "Pattern mining of multichannel sEMG for tremor classification," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 12, pp. 2795–2805, Dec. 2010.
[9] S. H. Roy, B. T. Cole, L. D. Gilmore, C. J. De Luca, and S. H. Nawab, "High resolution tracking of motor disorders in Parkinson's disease during unconstrained activity," *Mov. Disord.*, Mar. 20, 2013.
[10] M. Hoehn and M. Yahr, "Parkinsonism: Onset, progression and mortality," *Neurology*, vol. 17, no. 5, pp. 427–442, May 1967.
[11] E. Wan, "Discrete time neural networks," *J. Appl. Intell.*, vol. 3, pp. 91–105, 1993.

[12] L. Rabiner, "A tutorial on hidden Markov models and selected application in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.
[13] C. J. De Luca, "The use of surface electromyography in biomechanics," *J. Appl. Biomechan.*, vol. 13, pp. 135–163, 1997.
[14] B. T. Cole, S. H. Roy, C. J. De Luca, and S. H. Nawab, "Dynamic neural network detection of tremor and dyskinesia from wearable sensor data," presented at the 32nd Ann. Int. Conf. IEEE EMBS, Buenos Aires, Argentina, Sep. 1–4, 2010.
[15] S. Fahn *et al.*, "Unified Parkinson's disease rating scale," in *Recent Developments in Parkinson's Disease*, S. Fahn, C. D. Marsden, D. B. Calne, and M. Goldstein, Eds. Florham Park, NJ: Macmillan, 1987, pp. 153–163.
[16] M. R. Munetz and S. Benjamin, "How to examine patients using the abnormal involuntary movement scale," *Hospital Commun. Psychiatry*, vol. 39, no. 11, pp. 1172–1177, Nov. 1988.

Bryan T. Cole (S'10–M'12) was born in Warwick, R.I., USA, in 1985. He received the B.S.E. degree in electrical engineering and mathematics from Tulane University, New Orleans, LA, USA, in 2007, and the M.S. and Ph.D. degrees in electrical engineering from Boston University, Boston, MA, USA, in 2011 and 2012, respectively.

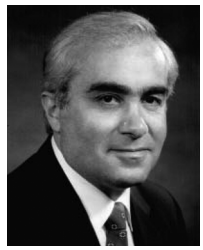
He is currently employed as a Research Engineer at Delsys, Inc., Natick, MA, USA. He was previously employed by Dräger Medical Systems and worked as a student in the Boston University NeuroMuscular Research Center. His current research interests include biomedical signal processing, dynamic machine learning algorithms, and wearable sensor systems.

Dr. Cole is a member of the Louisiana Beta chapter of Tau Beta Pi. He received a Dean's Graduate Fellowship from Boston University and the C. W. Ricker Award for Academic Excellence in Electrical Engineering from Tulane University.

Serge H. Roy received the B.S. degree from New York University, New York, USA, in 1975, and the M.S. and Sc.D. degrees from Boston University, Boston, MA, USA, 1981 and 1992, respectively.

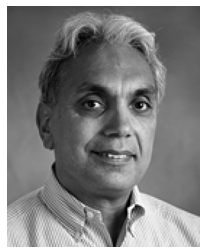
He is currently a Research Professor at Boston University NeuroMuscular Research Center and Sargent College of Health and Rehabilitation Sciences, Boston, MA, USA. His research interests are directed at assessing neuromuscular disorders using wearable sensors.

Dr. Roy is the recipient of the Elizabeth C. Adams Award (New York University) and two group achievement awards from NASA. He was elected as a Fellow to the American Institute for Medical and Biological Engineering (AIMBE) and to the International Society of Electrophysiology and Kinesiology (ISEK).



Carlo J. De Luca (LF'09) is the Founding Director of the Neuro Muscular Research Center at Boston University, Boston, MA, USA, and the Founding President and CEO of Delsys Inc., Natick, MA, USA. He holds the following positions: Professor of Biomedical Engineering, Professor of Electrical and Computer Engineering, Research Professor of Neurology, and Professor of Physical Therapy. He served as Dean of the College of Engineering from 1988 to 1990. He served a four year term on the National Advisory Board of NIH/NIBIB. He has 160 peer reviewed scientific publications, one book, over 21 000 citations, and 16 patents.

Dr. De Luca is a Founding Fellow of the American Institute for Medical and Biological Engineering (AIMBE) and a Founding Fellow of The Biomedical Engineering Society (BMES). He is the recipient of several scientific awards.



S. Hamid Nawab (SM'14) received the S.B., S.M., and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 1977, 1979, and 1982, respectively.

He joined the faculty of Boston University, Boston, MA, USA, in 1985, where he is currently Professor of Electrical and Computer Engineering. He has also held visiting professorships in Electrical Engineering at the Massachusetts Institute of Technology and in Computer Science at University of Massachusetts at Amherst. He has published over 100 research articles, including an award winning journal article on "Direction determination of wideband signals" in the IEEE TRANSACTIONS ON SIGNAL PROCESSING. His publications span a variety of contributions to core signal processing technologies, including short-time Fourier transforms, fast algorithms, approximate processing, integrated processing and understanding of signals, and machine learning, He has also published articles on applied signal processing for speech, audio, and biomedical applications. and has contributed eight chapters in DSP and DSP-related books. He is co-author of the textbook "Signals and Systems" (Prentice Hall, 1997). In addition, he is co-editor of the book "Symbolic and Knowledge Based Signal Processing" (Prentice Hall, 1992) on pioneering contributions at the intersection of DSP and artificial intelligence technologies.

Dr. Nawab was elected a Fellow of the American Institute of Medical and Biological Engineering in recognition of his contributions to the analysis of complex biosignals for speech, auditory, and neuromuscular systems, in 2006. As an educator, he has received several teaching awards