

User Training for Pattern Recognition-Based Myoelectric Prostheses: Improving Phantom Limb Movement Consistency and Distinguishability

Michael A. Powell, *Student Member, IEEE*, Rahul R. Kaliki, *Member, IEEE*, and Nitish V. Thakor, *Fellow, IEEE*

Abstract—We assessed the ability of four transradial amputees to control a virtual prosthesis capable of nine classes of movement both before and after a two-week training period. Subjects attended eight one-on-one training sessions that focused on improving the consistency and distinguishability of their hand and wrist movements using visual biofeedback from a virtual prosthesis. The virtual environment facilitated the precise quantification of three prosthesis control measures. During a final evaluation, the subject population saw an average increase in movement completion percentage from 70.8% to 99.0%, an average improvement in normalized movement completion time from 1.47 to 1.13, and an average increase in movement classifier accuracy from 77.5% to 94.4% ($p < 0.001$). Additionally, all four subjects were reevaluated after eight elapsed hours without retraining the classifier, and all subjects demonstrated minimal decreases in performance. Our analysis of the underlying sources of improvement for each subject examined the sizes and separation of high-dimensional data clusters and revealed that each subject formed a unique and effective strategy for improving the consistency and/or distinguishability of his or her phantom limb movements. This is the first longitudinal study designed to examine the effects of user training in the implementation of pattern recognition-based myoelectric prostheses.

Index Terms—Electromyography (EMG), motor learning, pattern recognition, prosthetics, rehabilitation, therapy.

I. INTRODUCTION

MYOELECTRIC prosthesis control has been accomplished with threshold-based direct and proportional control methods for several decades [1]. In direct control, a single electrode is responsible for monitoring muscle activity at a particular location in order to control a single prosthesis action. Proportional control is an extension of direct control in which the prosthesis action can be performed with variable velocity based on the amplitude of the detected electromyography (EMG) signal. While these methods may be adequate for a single-degree-of-freedom prosthesis, the increasing capabilities of commercially available prosthetic hands and wrist

devices cannot be intuitively controlled using the direct and proportional control paradigms. Many researchers have turned to pattern recognition as a means of analyzing multi-channel myoelectric signals in order to provide more intuitive control of advanced prostheses.

A. Pattern Recognition for Myoelectric Prostheses

Pattern recognition approaches to myoelectric prosthesis control vary widely in their implementation, but the set of problems addressed by classification algorithms remains relatively fixed. Myoelectric signals are processed, signal features are extracted, and classification decisions are made. In pattern recognition of myoelectric signals, misclassification can arise from changes in user effort or intention as well as one of many external factors. Such factors include motion artifact, electrode shift or lift-off, variable limb position, muscle fatigue, and electrode/skin contact [2]–[4]. We can avoid some of these external factors by conducting a controlled experiment in a virtual environment in order to better associate classification error with the subject's muscle activity (Fig. 1). Using a virtual prosthesis capable of every natural arm movement allows subjects to practice movements that no commercially available prosthesis can currently perform. A reasonable goal for classifier reliability is to provide predictable, uninterrupted use of a pattern recognition-controlled prosthesis that spans a standard workday lasting 8 h. Extending the reliability of a classifier might be possible through adaptive classification methods [5], but we investigate here whether or not prosthesis user training as defined in [1] can by itself play a significant role in extending the reliability of a static classifier.

Most modern classification techniques are capable of achieving impressive and nearly equivalent results when using the same robust features [6]. Linear discriminant analysis (LDA) is among these high-performing techniques, and its ease of implementation and low computational overhead make it one of the more attractive choices for real-time, closed loop models and embedded devices [7]. In this study, we use an LDA classifier to better associate improvements in subject performance with the effects of subject training rather than the sophistication of the classifier.

B. Role of Subject Training in Classifier Performance

The central question this study seeks to answer is whether an individual can significantly improve his or her real-time control of a pattern recognition-based virtual prosthesis through repeated user training. Numerous EMG pattern recognition

Manuscript received February 17, 2013; revised May 28, 2013 and August 01, 2013; accepted August 21, 2013. Date of publication October 07, 2013; date of current version April 28, 2014. This work was supported in part by the National Institutes of Health under Grant 1R44NS065495-01 and Grant 1R43HD072668-01 and in part by the General Omar N. Bradley Fellowship.

M. A. Powell and N. V. Thakor are with the Department of Biomedical Engineering, The Johns Hopkins University, Baltimore, MD 21218 USA (e-mail: mpowell35@jhu.edu).

R. R. Kaliki is with Infinite Biomedical Technologies, Baltimore, MD 21218 USA.

Digital Object Identifier 10.1109/TNSRE.2013.2279737

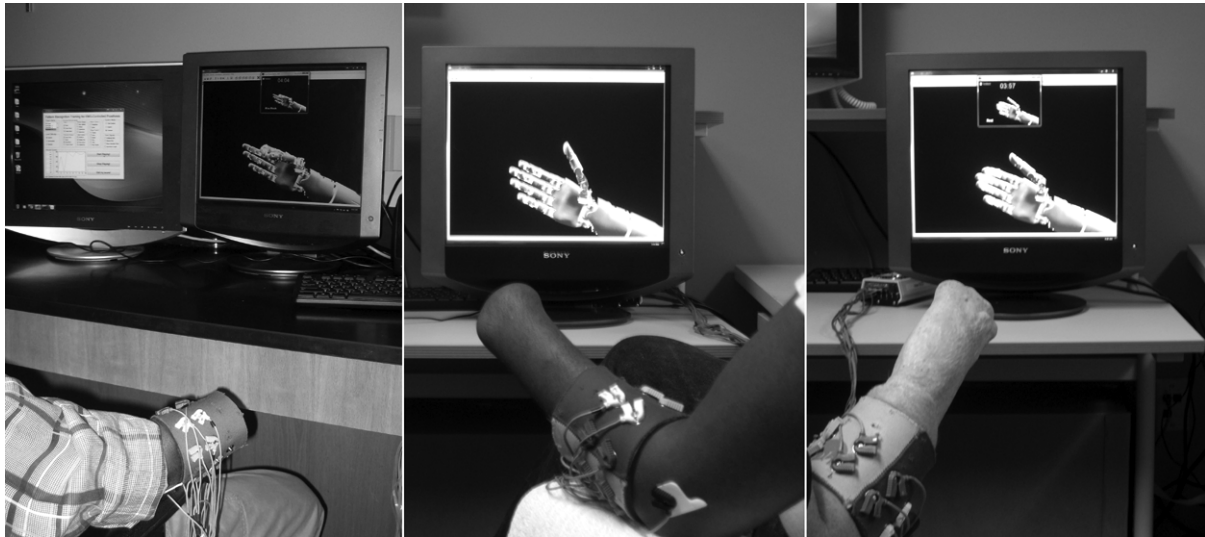


Fig. 1. Two unilateral transradial amputees (one pictured at left) and two unilateral wrist disarticulation amputees (center, right) learned to operate a virtual prosthesis controlled by pattern recognition. Subjects wore a silicone cuff embedded with eight electrode channels. Subjects conducted an initial familiarization and evaluation session, eight one-on-one training sessions, and a final evaluation session. A visual cue was provided above the virtual prosthesis (right image) to direct the subject to move the virtual prosthesis to the indicated hand and wrist position.

experiments have focused on algorithmic variations and novel classification techniques [8]–[15], but to date no other longitudinal study demonstrating subjects' ability to learn pattern recognition control has been published. Classifiers are often evaluated using able-bodied and amputee subjects who provide myoelectric signal data over a period lasting at most a couple of hours. Comparing one study to another can be difficult due to small sample sizes and the uniqueness of each amputee participant. There is perhaps even greater significance in the amount and type of feedback provided to study participants. Many classifiers are evaluated offline and do not provide the subject an opportunity to learn better control. In this wide range of pattern recognition experiments, the subjects may be completely naive or perhaps routine research volunteers with extensive exposure to EMG pattern recognition research. Short experiments with naive subjects are not reflective of the targeted prosthesis-using population. Amputees, however, interact daily with a prosthesis, so demonstrating that acceptable performance is achievable through user training shifts the focus of pattern recognition research from classifier sophistication to the development of patient training protocols and the implementation of pattern recognition-based prostheses. Studying user training also highlights the significance of testing advances in algorithms and control schemes using well-trained subjects. Assessing the effects of external factors in an untrained subject would group both the subject's inconsistencies and the variable of interest into a single, inseparable joint effect. While hardware and environmental factors will certainly limit the effectiveness of a pattern recognition-based prosthesis, we must first address the source of the greatest variability: the user.

II. METHODS

A. Equipment Setup

With no skin preparation, each transradial amputee subject wore a compressive silicone cuff on the muscle belly of his or

her residual forearm. The cuff was embedded with 3/8-in-diameter stainless steel dome electrodes with snap connectors (Liberating Technologies Inc., Holliston, MA). The cuff was equipped with 16 total electrodes arranged in eight bipolar pairs with 2-cm spacing. All electrode pairs were uniformly spaced circumferentially around the silicone cuff. The electrodes were aligned to run parallel to the longitudinal axis of the subject's arm (Fig. 1). An additional Norotrode 20 electrode (Myotronics Inc., Kent, WA, USA) was placed on the subject's olecranon and was connected as a reference electrode. The electrode pairs were differentially connected to Otto Bock EMG amplifiers (MYOBOCK, Otto Bock Healthcare, Minneapolis, MN, USA) modified for remote amplification. Amplifier outputs were connected to a desktop computer through a 16-bit data acquisition card (PCI-6040E, National Instruments, Austin, TX, USA) via a custom-made electrical isolation board and an I/O connector block (SCC-68, National Instruments, Austin, TX, USA). EMG signals were acquired at a sampling rate of 1 kHz.

During training sessions, subjects faced a computer screen and were presented visual cues that showed a robotic prosthesis in various hand and wrist positions alongside a corresponding textual cue (Fig. 1). The nine total movements included the following: hand at rest (R), hand open (HO), hand closed (HC), index finger point (IP), fine pinch (PI), forearm pronation (PR), forearm supination (FS), wrist flexion (WF), and wrist extension (WE). The virtual prosthesis was a model of the Modular Prosthetic Limb as constructed at the Johns Hopkins University Applied Physics Laboratory [16] and was visualized using the Musculo-Skeletal Modeling Software package known as MSMS [17]. This virtual prosthesis was animated to perform decoded movements by moving each controllable joint toward a target joint angle associated with the decoded movement. The movement rates of each joint were fixed for this experiment and approximated the movement rates of commercially available prostheses.

B. Classification Method

EMG signals were preprocessed using a 30–300 Hz third-order Butterworth filter, and the features were calculated for 200-ms analysis windows with a 175-ms overlap between successive windows. The LDA classifier utilized three extracted features calculated for each of eight channels of acquired EMG data. The features included the mean of the absolute value, the waveform length, and the variance [18].

C. Experiment Protocol

Four transradial amputees each attended 10 1-h sessions over the course of three weeks. The study consisted of a preliminary testing day, eight training days, and a final testing day. Only transradial participants were recruited for this study. All study candidates reported manageable levels of phantom limb pain. Had any interested candidate reported severe and persistent phantom limb pain, he or she would have been excluded due to the anticipated fatigue and pain such a subject would likely experience while repeatedly making residual and phantom limb movements.

1) *Preliminary Pattern Recognition Testing (Day 1 of 10)*: The purpose of the first day was to collect preliminary EMG data and to establish a baseline of performance prior to extensive user training. Once the electrodes were connected, the subject was asked to perform the following three tasks.

1) Train

The subject trained a movement classification algorithm by using his or her residual limb to mimic cues on a computer screen for the duration of each cue (Fig. 1). Each presented cue was visible for 4 s and then turned red for a final fifth second to alert the subject to the pending cue change. All movements were shown in a random order three times each for a total time of 2 min and 15 s. Training data was recorded during this phase in order to construct a classifier.

2) Explore

The subject explored the virtual prosthesis environment by attempting all trained movements and watching how the trained virtual prosthesis reacted to the subject's muscle signals. The goal of this exploration was to determine an optimal phantom limb position that reliably produced the desired virtual prosthesis action.

3) Evaluate

A study team member evaluated the subject's control of the virtual prosthesis by randomly presenting cues for the trained movements and measuring various indicators of subject performance. Each presented cue was visible for 6 s and then turned red for a final seventh second to alert the subject to the pending cue change. Subjects were instructed to attempt to achieve and hold the desired virtual prosthesis position for the duration of the cue. The cued positions were final positions such as a fully opened or closed hand; there were no intermediate positions. Holding the position required the subject to continuously provide muscle signals corresponding to the cued movement. Relaxing caused the virtual prosthesis to return to a rest position.

The preliminary recording consisted of a 1-h familiarization and evaluation session. Before obtaining the first test set, subjects became familiar with the task by performing the above steps several times with smaller movement subsets.

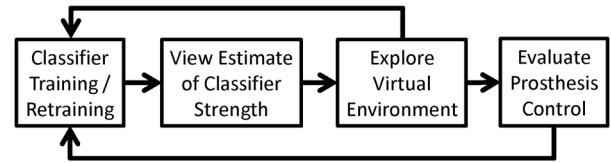


Fig. 2. Depicted here is the user training session workflow. A set of desirable movements was selected, and then the subject provided initial training data for classifier construction. The subject could then view an estimate of the classifier strength as depicted by a confusion matrix (see Fig. 5). This information was considered as the subject explored the virtual environment and attempted to control the virtual prosthesis to complete all trained movements. If the constructed classifier was unreliable, immediate modification and retraining was required. The goal of user training sessions was to reach the final step where an evaluation took place to measure progress and reinforce any learning that occurred during the user training session.

2) *Pattern Recognition User Training (Days 2–9 of 10)*: During user training sessions, the subject was coached by a study team member regarding the way in which pattern recognition algorithms decode movements. The study team member guided the participant through a progressive training plan that added additional movements as the subject became more proficient at controlling the virtual prosthesis. This progression was intended to take most of the three-week study period. The training sessions resembled the preliminary recording period. Subjects trained a movement classifier, observed the classifier's strength, and learned how to make it more reliable. The daily training sessions consisted of two or three iterations of the workflow depicted in Fig. 2. Through iterative refinement of each subject's motions during user training, we hypothesized that subjects would demonstrate improved control of the virtual prosthesis.

3) *Final Pattern Recognition Testing (Day 10 of 10)*: The final testing day at the end of the user training period resembled the first day of the study. The subjects tested the same set of movements to observe any improvement gained in virtual prosthesis control. The subjects tested once in the morning and then a second time 8 h later. The subjects wore the disconnected electrode cuff between sessions in order to test the stability of the movement classifier after one elapsed 8-h period. Subjects were instructed to complete all normal daily activities while being cautious not to shift the cuff in any way; subjects did not wear a prosthesis during this period. Marks were placed on each subject's residual limb to ensure the cuff had not moved. If obvious cuff displacement occurred, the cuff was adjusted prior to the second session to align it with the initial position markings (affected only Subject 3). The movement classifier was not retrained before evaluation during the second session in order to simulate testing the workday reliability of a static classifier.

D. Training Techniques

Subjects entered the study with varying abilities to perceive and control their phantom limbs based largely on the nature of each subject's amputation and the elapsed time since the amputation (see Table I for subject demographics). With each subject we first focused on movement consistency and then transitioned our efforts to increasing movement distinguishability. All of the study participants were unilateral amputees, which allowed each one to gain valuable insights by first training with

TABLE I
STUDY PARTICIPANT DEMOGRAPHICS

Subject	Sex	Age	Age at Amputation	Congenital or Traumatic	Level of Amputation	Prosthesis Used Daily
1	M	61	55	Traumatic	Short Transradial	Myoelectric and Body-Powered
2	F	38	28	Traumatic	Wrist Disarticulation	Body-Powered or No Prosthesis
3	M	56	6 mo	Traumatic	Wrist Disarticulation	Myoelectric and Body-Powered
4	M	51	11	Traumatic	Short Transradial	Myoelectric

his or her intact limb. Training with the intact limb exposed the subjects to the importance of muscle pattern consistency as well as the subtle differences that may cause the classifier to choose one movement over another. A more detailed explanation of our user training techniques appears in [19].

When consistency alone could not produce a reliably decoded movement, the confusion was typically found in one or two pairs of movements. In this circumstance, small modifications to one or more movements were necessary to better distinguish one movement from another. This forced the trainer and subject to come up with a unique, personalized focus for every movement. We adhered to two guidelines to ensure the resulting control was desirable.

- First, we asked the subject to describe the specifics of what he or she was doing to accomplish the movement. This ensured that the subject was aware of the technique that was producing success and could repeat it.
- Second, we required that the phantom limb movement be related to the desired movement. Natural, intuitive control should not be unnecessarily coupled with unrelated movements simply because they are easily distinguished.

The trainer and amputee focused on developing natural, intuitive control that the amputee could feel and confidently employ. In the case where a subject had no perception of a phantom limb, he or she first performed the desired motions with his or her intact limb and then attempted to mimic the perceived muscle contractions of the intact limb using the residual limb.

E. Quantifying Virtual Prosthesis Control

Classification algorithms are often evaluated by comparing all decisions made to the known class of the evaluated data. It is important to ensure that the metrics used in our application reflect the reliability of a prosthesis to the end user. For example, an instantaneous but fleeting correct decision has no real significance. Similarly, the effects of spurious misclassifications may be attenuated by a majority voting [18] or velocity ramp scheme [20]. Only a consistent, correct decision stream will see a prosthesis reach the desired end state. The user is primarily concerned with whether or not a posture can be achieved and whether or not that posture can be maintained. Tests and corresponding metrics that evaluate true prosthesis control have been developed by Kuiken, *et al.* with the Motion Test in [21] and Simon, *et al.* with the Target Achievement Control Test in [22]. We quantified our assessment of each subject using three modified metrics stemming from these published tests.

1) *Movement Completion Percentage (MCP)*: MCP reflects the percentage of movements a subject completed within the individual movement cue periods. Each cue period started with the

subject and virtual prosthesis at rest. The virtual prosthesis at all times moved toward the current predicted movement, which did not always match the cued movement. A movement was considered complete when the user successfully moved the virtual prosthesis through a full range of motion such that all model joint angles were within a small tolerance of the target position as indicated by a visual cue. The allowable tolerance was 1.125° for each joint angle, which corresponded to the maximum angular displacement each joint was allowed to move for each 25-ms time step. Higher movement completion percentages indicate better performance, and the maximum score using this metric is 100%.

2) *Average Movement Completion Time (MCT)*: All timed movements were initiated from a rest position in order to uniformly measure the total time required for the user to move the virtual prosthesis to the desired end state. The timer started at the instant the user demonstrated a muscle pattern that yielded a movement decision other than “rest,” regardless of whether or not it was the correct movement. For movements that were eventually completed, this time measurement was then compared to the optimal movement time as established by the model’s movement rate parameters. Since various movements require different lengths of time, we normalized the MCT and then averaged the normalized MCTs across all movements. A lower normalized MCT is better, and the best possible score is 1.

3) *Post-Selection Classification Accuracy (PSCA)*: A final attribute of user control over the virtual prosthesis is the stability of the decision stream after a movement has been correctly identified. We considered all classification decisions made after the first correct decision through the end of the movement cue. This eliminated the period considered to be reaction time and movement onset or transition time. Higher PSCA indicates better stability, and the maximum score using this metric is 100%. PSCA further refines the MCP and MCT by now showing how stably a subject could maintain the desired position upon reaching it. Note that this required the subject to continue providing EMG signals that corresponded to the provided movement cue. In practice a subject may prefer to use the “rest” class to hold a prosthesis in a fixed position, but for subject training purposes, we required the subjects to hold the desired position for the duration of each cue.

In summary, this three-tiered approach focused first on whether a movement could be achieved at all. Second, we measured the efficiency or speed with which a movement could be completed. Finally, we measured the stability of the completed movements to assess the subject’s overall ability to control a virtual prosthesis.

F. Statistical Analysis

The performance metrics we use to evaluate virtual prosthesis control are best evaluated for significance using bootstrap distributions. Each metric has a limit corresponding to perfect virtual prosthesis control, and our expectation that subjects will approach these limits implies that the associated distributions would be skewed. Using a bootstrap distribution allows us to compute significance for a relatively small sample size in which assumptions of normality would be inappropriate [23]. Our test for significance for each reported metric is implemented in MATLAB and utilizes repeated resampling with replacement ($n = 1000$) to compute two bootstrap distributions approximating two population means (i.e., MCP_{Day1} versus MCP_{Day10}). We then compute a statistic that compares the two distributions by subtracting one full set of random samples from another full set of random samples ($MCP_{Day10} - MCP_{Day1}$). A p-value is computed by counting the frequency with which the computed statistic does not support the hypothesis. For example, we hypothesize that $MCP_{Day10} > MCP_{Day1}$. The corresponding p-value for this hypothesis is $p = \text{count}(MCP_{Day10} - MCP_{Day1} < 0)/n$.

G. Measuring Class Separability

There were two primary goals for this study: 1) improve a subject's virtual prosthesis control, and 2) assess the reliability of a trained classifier after 8 h. Additionally, we hoped to understand which strategies amputees employed that led to improved pattern recognition control. Improvement in prosthesis control is conceptually equivalent to increasing the separability of movement classes in the feature space of an LDA classifier. When a subject encountered movement pairs that were not decoded with great accuracy, we focused on increasing the separability of those specific classes. This necessarily requires some measure of separability to assess whether the user training was effective.

Several measures of class separability are proposed in [24] that address the distance between class distributions, and these include the Euclidean and Mahalanobis distances as well as Fisher's linear discriminant ratio. Other proposed measures include the repeatability index, separability index, and cluster size approximation using the mean of the semi-principal axes defined in [25]. Fisher's linear discriminant ratio compares between-class scatter to within-class scatter [26], and this ratio is the foundation of linear discriminant analysis. To present this ratio, we define total scatter (Σ_t), within-class scatter (Σ_w), and between-class scatter (Σ_b)

$$\begin{aligned}\Sigma_t &= \sum_{i=1}^c \left[\sum_{j=1}^{n_i} (x_{ij} - \mu)(x_{ij} - \mu)^T \right] \\ \Sigma_w &= \sum_{i=1}^c \left[\sum_{j=1}^{n_i} (x_{ij} - \mu_i)(x_{ij} - \mu_i)^T \right] \\ \Sigma_b &= \sum_{i=1}^c n_i (\mu_i - \mu)(\mu_i - \mu)^T \\ &= \Sigma_t - \Sigma_w\end{aligned}\quad (1)$$

where c is the total number of movement classes, n_i is the number of samples in class i , x_{ij} is the j th individual sample (feature vector) from class i , μ is the mean feature vector for all classes, and μ_i is the mean feature vector for class i .

We now calculate a separability measure S to describe the ratio of between-class scatter to within-class scatter

$$S = \frac{\text{tr}(\Sigma_b)}{\text{tr}(\Sigma_w)}. \quad (2)$$

An increasing value of S describes a more separable data set. This formula drives the user training techniques we employed with the subjects in this study. As observed by Bunderson, *et al.* in [25], subject training could have two objectives: increase between-class scatter and/or reduce within-class scatter.

We hypothesized that practice alone would contribute to reduced within-class scatter, especially with real-time feedback informing the subject of correct and incorrect movement classification decisions. Like any motor task, focused repetition was expected to build consistency, thus reducing the variability that a naive subject would likely and unintentionally demonstrate. We also expected that consistency alone might not greatly improve movement discrimination for some subjects. Prior experience suggests that visibly different movements are sometimes quite similar in the feature space, making them difficult to reliably decode. For pairs of movements exhibiting this quality, intentional changes to phantom hand or wrist movements were required to separate movement pairs in the feature space [19].

H. Sources of Improvement

To address the question of whether user training should focus on decreasing within-class scatter, increasing between-class scatter, or perhaps both, we consider each subject individually using various measures of their exhibited movement class distributions. These measures include the Fisher linear discriminant ratio, Euclidean distance, Mahalanobis distance, and the mean of the semi-principal axes as an approximation of cluster size.

1) *Separability (S)*: Separability is a pairwise metric used to validate user training success that stems from the Fisher linear discriminant ratio. The equation is developed in (2). Separability is the primary indicator of expected performance.

2) *Euclidean Distance (D_E)*: Euclidean distance is another pairwise metric used to assess changes in the arrangement of class clusters in high dimensional space. The Euclidean distance is calculated from centroid to centroid of every movement pair using the following equation:

$$D_E(a, b) = \sqrt{\sum_{i=1}^n (b_i - a_i)^2} \quad (3)$$

where i represents one of the n dimensions that describe the location of the centroids for classes a and b .

3) *Mahalanobis Distance (D_M)*: Mahalanobis distance as calculated here measures the distance between two classes a and b with consideration given to the variance of each class in the space between the centroids. It is a refinement of Euclidean distance that incorporates the size and shape of class distributions.

It is often used to calculate the variance-scaled distance from a distribution to a point in space

$$D_M(\mu_a, \mu_b) = \sqrt{(\mu_b - \mu_a)^T \Sigma^{-1} (\mu_b - \mu_a)} \quad (4)$$

where μ_a and μ_b are the mean vectors for movement classes a and b , and Σ is the covariance matrix for all data points belonging to movement classes a and b .

4) *Mean of the Semi-Principal Axes (MSA)*: The mean of the semi-principal axes is a measure of movement class cluster size proposed in [25]. We consider changes in individual cluster size to represent changes in movement execution consistency. The cluster shape is considered to be a hyperellipsoid, and its size is approximated using the product of the singular values obtained through singular value decomposition

$$MSA(a) = \left(\prod_{i=1}^n a_i \right)^{\frac{1}{n}} \quad (5)$$

where a_i is the i th of n singular values obtained by decomposing a matrix of all data belonging to movement a .

III. RESULTS

In an IRB-approved study conducted at the National Rehabilitation Hospital, Washington, D.C., USA, four transradial amputees completed a 10-session pattern recognition study.

A. User Training Effect

Our initial objective was to gauge the effectiveness of our user training methods for improving a subject's pattern recognition control of a virtual prosthesis. We found significant improvement in movement completion percentage, movement completion time, and post-selection classification accuracy (Fig. 3). The following group means are reported along with the p-value resulting from the bootstrap distributions described in Section II-F. The mean movement completion percentage for the group of four amputees improved from 70.8% to 99.0% ($p < 0.001$). The mean normalized movement completion time for the group of four amputees improved from 1.47 to 1.13 ($p < 0.01$). For each subject we considered only those movements from Day 10 that were achieved at least once on Day 1 in order to make a fair comparison. All subjects were capable of achieving between seven and nine of the original nine movement classes on the first day. The mean post-selection classification accuracy for the group of four amputees improved from 77.5% to 94.4% ($p < 0.001$). With few individual exceptions, the group as a whole improved significantly in each measure of virtual prosthesis control.

B. User Training Effect on Day-Long Reliability

A second objective was to consider the effect of user training on the day-long reliability of a movement classifier (Fig. 4). We expected subjects to be able to maintain or only negligibly decline in performance over 8 h. The mean movement completion percentage for the group of four amputees declined from 99.0% to 95.8%, although the difference was not statistically significant ($p = 0.116$). The mean normalized movement completion time for the group of four amputees increased from

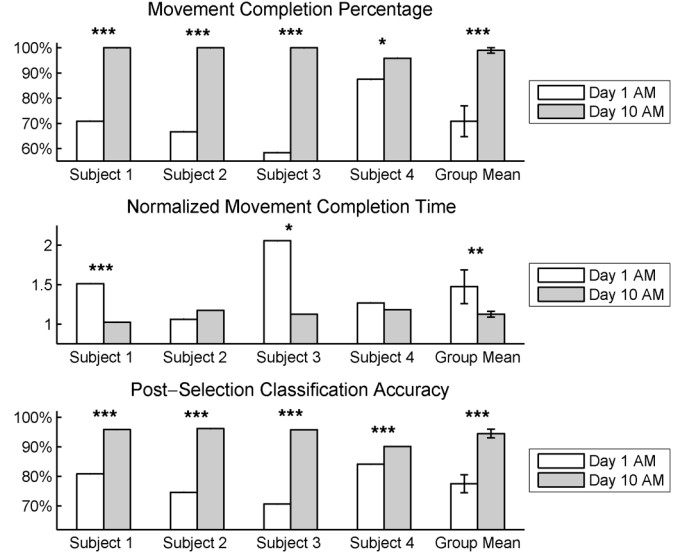


Fig. 3. Depicted here are the performance measures for four transradial amputees before and after a 10-session user training study. Subjects were instructed to complete virtual prosthesis movements in response to visual cues. Control of the virtual prosthesis was measured by the percentage of tasked movements a subject completed (top), the time required to complete the movements (middle), and the stability of the classification decision stream following correct movement selection (bottom). Asterisks indicate significance (*: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$).

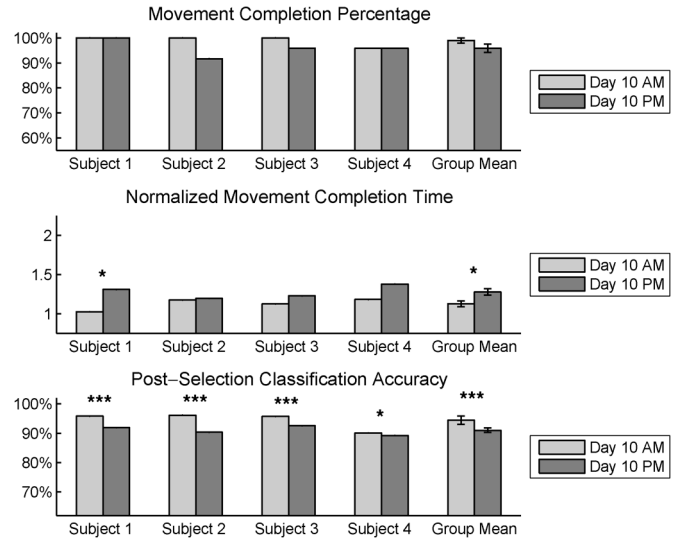


Fig. 4. Depicted here are the same three performance measures as seen in Fig. 3 with 8 h elapsed between sessions. Subjects trained and evaluated a movement classifier during the final morning of the study. Subjects wore the electrode cuff for 8 h and were tested again without retraining the movement classifier. Subject 3 required a small cuff realignment due to obvious displacement of the cuff (approx. 1.5-cm shift distally) during the 8 h between sessions. Asterisks indicate significance (*: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$).

1.13 to 1.27, which proved significant ($p < 0.05$). All movements were considered for all subjects because all subjects successfully achieved at least one of each movement class in both sessions. The post-selection classification accuracy for the group of four amputees declined from 94.4% to 91.0% ($p < 0.001$). The group as a whole experienced insignificant changes in movement completion percentage after 8 elapsed

TABLE II
STUDY PARTICIPANT LEARNING TIMELINE

Subject	Day 1	Study Days Required to Reach				Day 10
	MCP	75%	90%	95%	100%	MCP
1	71%	2	2	2	2	100%
2	67%	3	3	4	8	100%
3	58%	7	7	7	7	100%
4	88%	1	2	2	6	96%

Each subject's end-of-study evaluation score was independent of previous performance, so it was possible for a subject to score lower on the final day than the same subject may have scored on an earlier day in the study. This was the case for Subject 4.

TABLE III
TRAINING AND EVALUATION MOVEMENTS

Abbreviation	Hand / Wrist Movement
R	Rest
HO	Hand Open
HC	Hand Close
IP	Index Point
PI	Fine Pinch
PR	Forearm Pronation
FS	Forearm Supination
WF	Wrist Flexion
WE	Wrist Extension

hours, but they demonstrated a statistically significant increase in movement completion time and decline in post-selection classification accuracy. This decline in PSCA should affect a subject's ability to maintain a classified movement, but in practice this decline was small enough that the sufficiently high residual classification accuracy made any emergence of unreliability undetectable.

C. Timeline for Improvement

There is no single pace that can describe how each person progressed through this user training study. Table II shows several performance milestones and the corresponding days on which each subject reached each milestone. Subject 1 needed just one day of training to master the pattern recognition concepts and implement them. Subjects 2 and 4 required more time, and their progressions were gradual. Subject 3 made no significant improvements until the seventh day, but the concepts and practiced techniques were then understood completely and led to near-perfect performance for the remainder of the study. Each individual's day-to-day performance varied while trying different techniques, but prior achievement of MCP performance milestones gave each participant the ability to revert back to a proven technique.

D. Individual Subject Separability Analyses

The confusion matrices in Figs. 5–8 represent subject performance on the first and last day of the study and are useful for identifying specific movement pairs that experienced a noticeable change in classification accuracy. We set a minimum threshold of a 1% change in the accuracy represented by a diagonal element or the error represented by an off-diagonal element. If any movement pair experienced such a change, that movement pair as well as the individual movements making up that pair are considered to be of interest in the forthcoming analysis. If no significant change occurred, we removed the associated movement pair from the analysis as any calculated changes had no effect on user control of the virtual prosthesis. This primarily occurred when the user had near-perfect discrimination of a movement pair both before and after user training.

In the additional measures presented in Figs. 5–8, all possible movement pairs are shown in the upper right portion of each matrix excluding the diagonal. White squares indicate an improvement in the associated metric after user training. Black squares indicate a decline in the associated metric after user training.

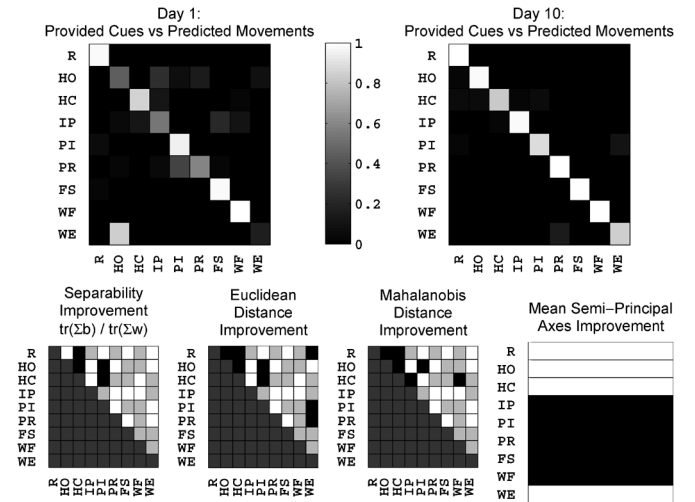


Fig. 5. Presented here are confusion matrices denoting Subject 1's Day 1 and Day 10 evaluation sessions. Additionally, four performance metrics are provided to explain the source of any positive or negative changes observed in the confusion matrices. As described in Section III-D, white squares signify improvement in the indicated metric. Black squares signify decline. Gray squares signify a lack of significant change in subject performance for the movements in question.

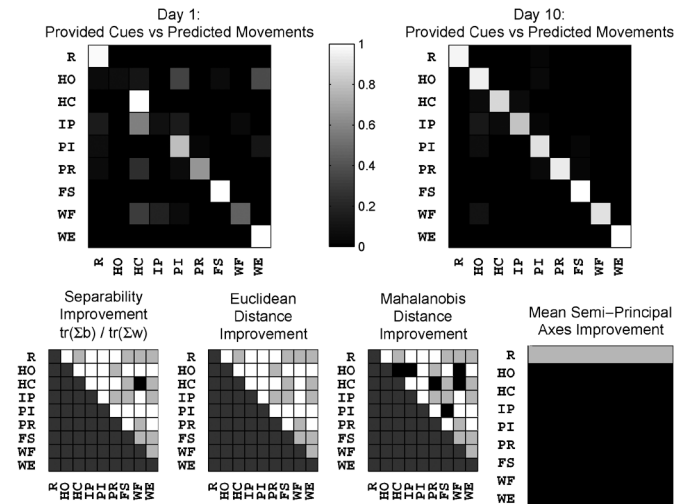


Fig. 6. Subject 2's confusion matrices and additional metrics are presented in the same manner as Fig. 5.

Gray squares indicate that negligible change ($\leq 1\%$) was reflected in the confusion matrices as a result of user training. All study participants trained and evaluated the full set of movements listed in Table III.

TABLE IV
SUBJECT-SPECIFIC CHANGES IN EFFORT

Subject	Average Final vs. Initial Movement Force Ratio	Strategy Interpretation
1	1.03 ± 0.14	Subject modified individual movement efforts without affecting overall effort.
2	1.12 ± 0.12	Subject modified individual movement efforts and increased overall effort.
3	0.86 ± 0.12	Subject modified individual movement efforts and decreased overall effort.
4	1.00 ± 0.02	Subject made negligible changes to individual movement efforts and overall effort.

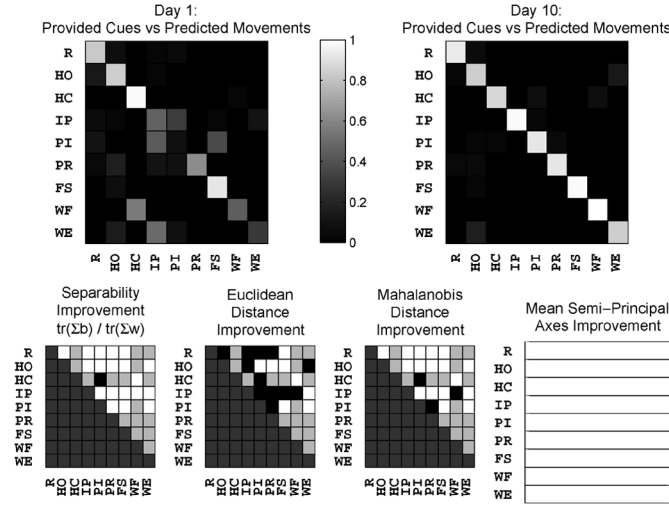


Fig. 7. Subject 3 was the only study participant who had no phantom limb sensation at the beginning of the study. While other subjects worked to refine their phantom limb positions during the study, Subject 3 learned to produce distinguishable muscle signal patterns by mimicking intact limb muscle contractions. Subject 3 achieved performance levels comparable to the other subjects despite lacking a phantom limb. Subject 3's confusion matrices and additional metrics are presented in the same manner as Fig. 5.

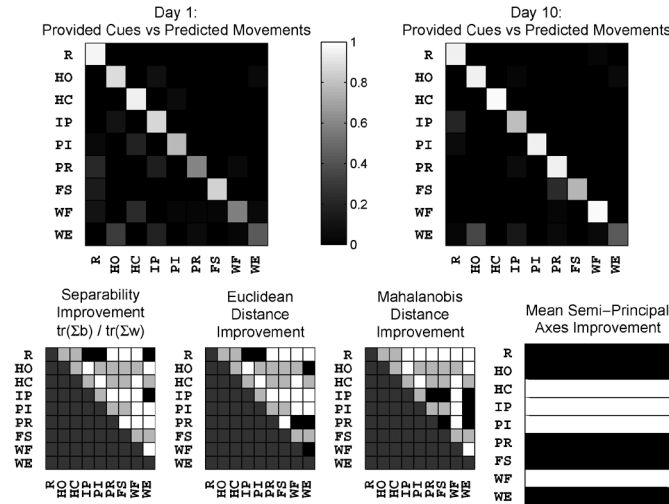


Fig. 8. Subject 4's confusion matrices and additional metrics are presented in the same manner as Fig. 5.

E. Muscle Contraction Strength as an Indicator of Effort

Some subjects were able to improve their movement pair discrimination while bringing classes closer together in the feature space. We consider that perhaps the subjects' improvement in

consistency allows for weaker muscle contractions to still generate reliable results. It is likely that these weaker muscle patterns would be nearer to one another in the feature space, perhaps all closer to the rest state than they were before extensive user training. Excluding the rest state, we examine the eight remaining movements for any increase or decrease in mean force used to generate the movement as a result of user training. We propose the following as a unitless measure of total force perceived by the eight electrode channels:

$$\mathbf{F}_i = \sum_j^n f_{i,j}^2 \quad (6)$$

where \mathbf{F}_i is the total muscle force for movement i , $f_{i,j}$ is the mean absolute value of all examples of movement i as recorded by channel j during the entire evaluation session.

Movement onset can often be characterized by a burst in EMG amplitude that exceeds the steady-state amplitude associated with the "hold" phase of a movement. One might expect that a subject's initial user training session might exhibit higher overall force due to the frequent corrections the subject might make to his or her phantom limb position. While searching for the correct phantom limb position may lead to multiple unintended EMG bursts in the same movement cue, Figs. 5–8 suggest that some subjects intentionally increased the force between the initial and final sessions. Trends emerge for each subject that tell us overall how movements changed over the course of user training. We observe in Table IV that the individual is quite capable of discovering his or her own strategy to become successful.

IV. DISCUSSION

A. Diversity of Subject Population

Table I highlights the differences between study participants' amputations and prosthesis usage. The only thing these subjects had in common was a lack of prior experience controlling a virtual prosthesis using pattern recognition. Their initial skill levels were quite different, and yet they obtained very similar levels of control by the end of the study. This suggests that long-term patient outcomes for pattern recognition-based prostheses cannot be accurately estimated from a single, initial testing session. Subject 3 had a particularly difficult time initially due to a complete absence of phantom limb sensation. Subject 3's left hand was amputated at the age of six months, and there was no memory or mental model of that hand to exercise when attempting virtual prosthesis movements. Despite this disadvantage, Subject 3 attained overall performance that was as good or better than any other subject. Subject 2 was the only

subject with no myoelectric prosthesis experience, and understandably Subject 2 struggled initially to accurately employ different muscle groups that had gone unused for a decade. While it did take Subjects 2 and 3 the longest to reach 100% MCP, they both eventually matched the skill demonstrated by Subjects 1 and 4. Thus, extensive user training allowed all subjects to reach comparable performance levels.

B. User Training Benefits for Day-Long Use

This study shows that a trained subject can provide consistent muscle patterns capable of being accurately decoded for a period of at least 8 h. This does not guarantee that variable socket forces, changing skin or environmental conditions, or excessive muscle fatigue will not affect classifier decisions; however, we have demonstrated that subjects are capable of producing consistent, distinguishable muscle patterns that do not require adaptive classification. Another benefit resulting from repeated user training in the virtual environment was the ability of each subject to produce a highly reliable training data set on the first try by the end of the study. Subjects were better able to respond promptly and accurately to training cues. The consistency and distinguishability present in each subject's movements would allow the subject to retrain a classifier with confidence if retraining were ever required. In this study, retraining is accomplished in less than 2.5 min; thus our approach would provide the subject a quick and reliable fix if a pattern recognition-based prosthesis experienced environmental changes extreme enough to make the original classifier unreliable.

C. Sources of Improvement

A surprising finding in the analysis of each subject's sources of improvement was the range of strategies unknowingly employed by various subjects. Subject 1's improvement came from a combination of increasing inter-class distances and reducing class cluster sizes (see Fig. 5). An example of increasing inter-class distances is the separability improvement between index point (IP) and forearm supination (FS). The separability of these two classes improved despite the increased cluster size of both movements. This is an example of increased distinguishability. An example of increased consistency is seen between rest (R) and wrist extension (WE). The reduced size of each class cluster led to increased separability despite the fact that the movement pair saw a decrease in Euclidean distance between class centroids. Subject 2's improvement can be exclusively attributed to making more distinguishable movements (see Fig. 6). Every single movement pair of interest saw an increase in inter-class Euclidean distance, and every class cluster saw an increase in cluster size. In all but one movement pair, the net result was increased separability. As suggested in Table IV, this assessment confirms that Subject 2's strategy was to move the classes farther apart. Subject 3's improvement came from reduced cluster sizes for every movement as well as a few increased inter-class distances (see Fig. 7). There are several examples where the Mahalanobis distance between classes increased despite a decrease in Euclidean distance as a result of the decreased cluster sizes. This suggests a greater reliance on consistency than on distinguishability. As suggested in Table IV, this assessment confirms

that Subject 3's strategy was to focus on consistency while exerting less effort. Subject 4's improvement was due to a balance of consistency and distinguishability efforts (see Fig. 8). The separability improvement between rest (R) and forearm pronation (PR) came from an increased inter-class distance despite the larger class clusters. Consistency helped reduce forearm pronation (PR) and wrist flexion (WF) confusion due to the smaller wrist flexion class cluster even though the classes moved closer together.

Based on the observed effectiveness of each subject's unique strategy, it is likely unnecessary and potentially counterproductive to promote a single strategy to a subject. The varying strategies each subject employed may largely relate to the amount of phantom limb control the subject felt. One subject may find it easier to vary force for increased distinguishability while another prefers more subtle variations in the individual digits of the phantom hand. Monitoring strategy choice might be possible using the various indicators of cluster sizes and separation during subject training, but it would be useful primarily to investigate a negative trend. For example, two movements may have increased in individual cluster size and moved closer together in the feature space. Increased cluster size may simply be an indicator of increased force used during the movement. A stronger muscle contraction should exhibit increased signal-dependent noise [27], thus enlarging the associated class cluster. Provided that this larger cluster moves further away in the feature space, this may still yield a positive result as observed in Fig. 6. Using this type of information in a subject training scenario is likely too abstract to be instructive, and relying instead on the confusion matrix produced at the end of each session provides a simple and adequate visual explanation of both overall performance and specific areas needing improvement.

D. The Role of Biofeedback

Because this virtual prosthesis used an LDA classifier and always moved toward the predicted hand position, incorrect classifications did not always look like the movement the subject was attempting. This can be a very confusing concept initially, but each subject learned this concept by exploring the boundaries in phantom limb space that separate movements like open hand and close hand. Subjects then saw the importance of focusing on achieving the correct movement rather than trying to understand why a particular incorrect movement was predicted.

Subjects often learn best when their focus is not specifically on the mechanics of their movement, but rather on some external object [28]. Controlling a virtual hand in real-time allows subjects to focus on an external object and assess the effects of their own motor movements. Movements may feel the same if the subject is focused strictly on the mechanics of a hand or wrist movement, but interpreting the feedback of a real-time movement decoding system allows the subject to discern the subtle differences in a way that translates into intuitive, subconscious control of additional degrees of freedom. This feedback is either confirmation that the desired muscle pattern was produced or proof that a different muscle pattern was presented.

The role of a biofeedback training system for pattern recognition control of a prosthesis is essential, and yet the supply of

feedback to the user must be rationed in a way that does not distract from the task. It is sometimes the case that the user's attention and efforts are more directed to addressing or responding to the feedback than to completing the intended task [29]. Biofeedback should be informative and perhaps motivating without becoming a distraction.

E. Future Work

Successful operation of a pattern recognition-based prosthesis will require focused user training for nearly all potential users. This in no way implies that only the most capable myoelectric prosthesis users are suitable candidates for pattern recognition-based prostheses. In fact, the candidate pool for pattern recognition-based prostheses is likely much larger than many researchers would expect. This transition may be more difficult or more time consuming for some amputees, but this study suggests that this transition can be successfully completed by a wide range of amputees. With a functional pattern recognition-based prosthesis, we can extend this work with well-trained subjects to study posture effects and other external factors. Functional user training with a prosthesis adds the realism of maintaining control of an object versus dropping it, and such experiments will truly quantify the effectiveness of both system training and user training. As this was a study primarily focused on improving class separability through user training measured with virtual prosthesis performance metrics, there is a need for task-based evaluation in a functional environment while performing activities of daily living.

V. CONCLUSION

Effective user training that utilizes real-time biofeedback is critical while learning a new technique for prosthesis control. Subjects in this study demonstrated that this type of user training not only leads to improved virtual prosthesis control, but it also makes classifier reliability for at least a standard workday a likely outcome for a trained user. As hardware and sophisticated algorithms work to address the external factors in pattern recognition performance, we must acknowledge that the effectiveness of a pattern recognition-based prosthesis ultimately lies in comprehensive user training. Through user training, amputees can learn to produce the consistent, distinguishable muscle patterns necessary for robust pattern recognition-based prosthesis control.

ACKNOWLEDGMENT

The authors would like to thank Dr. A. Dromerick and R. Holley of the Arm Amputee Program at the Center for Upper Limb Rehabilitation and Research (CULRR) at MedStar National Rehabilitation Hospital for hosting this study.

REFERENCES

- [1] A. Fougner, O. Stavdahl, P. Kyberd, Y. Losier, and P. Parker, "Control of upper limb prostheses: Terminology and proportional myoelectric control—A review," *IEEE Trans Neural Syst. Rehabil. Eng.*, vol. 20, no. 5, pp. 663–677, Sep. 2012.
- [2] A. Fougner, E. Scheme, A. Chan, K. Englehart, and Ø. Stavdahl, "Resolving the limb position effect in myoelectric pattern recognition," *IEEE Trans Neural Syst. Rehabil. Eng.*, vol. 19, no. 6, pp. 644–651, Dec. 2011.
- [3] L. Chen, Y. Geng, and G. Li, "Effect of upper-limb positions on motion pattern recognition using electromyography," in *Proc. 4th Int. Congr. Image Signal Process. Conf.*, 2011, vol. 1, pp. 139–142.
- [4] L. Hargrove, K. Englehart, and B. Hudgins, "A training strategy to reduce classification degradation due to electrode displacements in pattern recognition based myoelectric control," *Biomed. Signal Process.*, vol. 3, no. 2, pp. 175–180, 2008.
- [5] J. Sensinger, B. Lock, and T. Kuiken, "Adaptive pattern recognition of myoelectric signals: Exploration of conceptual framework and practical algorithms," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 17, no. 3, pp. 270–278, Jun. 2009.
- [6] E. Scheme and K. Englehart *et al.*, "Electromyogram pattern recognition for control of powered upper-limb prostheses: State of the art and challenges for clinical use," *J. Rehabil. Res. Develop.*, vol. 48, no. 6, p. 643, 2011.
- [7] A. Young, T. Hargrove, and L. Kuiken, "The effects of electrode size and orientation on the sensitivity of myoelectric pattern recognition systems to electrode shift," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 9, pp. 2537–2544, Sep. 2011.
- [8] K. Englehart, B. Hudgin, and P. Parker, "A wavelet-based continuous classification scheme for multifunction myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 3, pp. 302–311, Mar. 2001.
- [9] Y. Huang, K. Englehart, B. Hudgins, and A. Chan, "A Gaussian mixture model based classification scheme for myoelectric control of powered upper limb prostheses," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 11, pp. 1801–1811, Nov. 2005.
- [10] M. Lucas, A. Gaufriau, S. Pascual, C. Doncarli, and D. Farina, "Multi-channel surface EMG classification using support vector machines and signal-based wavelet optimization," *Biomed. Signal Process.*, vol. 3, no. 2, pp. 169–174, 2008.
- [11] C. Choi, S. Micera, J. Carpaneto, and J. Kim, "Development and quantitative performance evaluation of a noninvasive EMG computer interface," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 1, pp. 188–191, Jan. 2009.
- [12] L. Hargrove, E. Scheme, K. Englehart, and B. Hudgins, "Multiple binary classifications via linear discriminant analysis for improved controllability of a powered prosthesis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 18, no. 1, pp. 49–57, Feb. 2010.
- [13] F. Mattioli, E. Lamounier, A. Cardoso, A. Soares, and A. Andrade, "Classification of emg signals using artificial neural networks for virtual hand prosthesis control," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2011, pp. 7254–7257.
- [14] F. Chan, Y. Yang, F. Lam, Y. Zhang, and P. Parker, "Fuzzy EMG classification for prosthesis control," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 3, pp. 305–311, Sep. 2000.
- [15] A. Chan and K. Englehart, "Continuous myoelectric control for powered prostheses using hidden Markov models," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 1, pp. 121–124, Jan. 2005.
- [16] M. Johannes, J. Bigelow, J. Burck, S. Harshbarger, M. Kozlowski, and T. Van Doren, "An overview of the developmental process for the modular prosthetic limb," *Johns Hopkins APL Tech. Dig.*, vol. 30, no. 3, pp. 207–216, 2011.
- [17] R. Davoodi and G. Loeb, "Real-time animation software for customized training to use motor prosthetic systems," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 2, pp. 134–142, Mar. 2012.
- [18] K. Englehart and B. Hudgins, "A robust, real-time control scheme for multifunction myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 7, pp. 848–854, Jul. 2003.
- [19] M. Powell and N. Thakor, "A training strategy for learning pattern recognition control for myoelectric prostheses," *J. Prosthet. Orthot.*, vol. 25, no. 1, pp. 30–41, 2013.
- [20] A. Simon, L. Hargrove, B. Lock, and T. Kuiken, "A decision-based velocity ramp for minimizing the effect of misclassifications during real-time pattern recognition control," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 8, pp. 2360–2368, Aug. 2011.
- [21] T. Kuiken, G. Li, B. Lock, R. Lipschutz, L. Miller, K. Stubblefield, and K. Englehart, "Targeted muscle reinnervation for real-time myoelectric control of multifunction artificial arms," *JAMA*, vol. 301, no. 6, pp. 619–628, 2009.
- [22] A. Simon, L. Hargrove, B. Lock, and T. Kuiken, "Target achievement control test: Evaluating real-time myoelectric pattern-recognition control of multifunctional upper-limb prostheses," *J. Rehabil. Res. Develop.*, vol. 48, no. 6, pp. 619–28, 2011.
- [23] T. Hesterberg, D. S. Moore, S. Monaghan, A. Clipson, and R. Epstein, "Bootstrap methods and permutation tests," in *Introduction to the Practice of Statistics*. New York: W.H. Freeman, 2005, vol. 5, pp. 1–70.

- [24] I. Skrypnik, "Irrelevant features, class separability, and complexity of classification problems," in *Proc. 23rd IEEE Int. Conf. Tools Artif. Intell.*, 2011, pp. 998–1003.
- [25] N. Bunderson and T. Kuiken, "Quantification of feature space changes with experience during electromyogram pattern recognition control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 3, pp. 239–246, May 2012.
- [26] L. Wang, "Feature selection with kernel class separability," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 9, pp. 1534–1546, Sep. 2008.
- [27] B. Prilutsky, D. Ashley, L. VanHiel, L. Harley, J. Tidwell, and D. Backus, "Motor control and motor redundancy in the upper extremity: Implications for neurorehabilitation," *Top. Spinal Cord Inj. Rehabil.*, vol. 17, no. 1, pp. 7–15, 2011.
- [28] N. McNevin, G. Wulf, and C. Carlson, "Effects of attentional focus, self-control, and dyad training on motor learning: Implications for physical rehabilitation," *Phys. Ther.*, vol. 80, no. 4, pp. 373–385, 2000.
- [29] R. Schmidt and R. Bjork, "New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training," *Psychol. Sci.*, vol. 3, no. 4, pp. 207–217, 1992.



Michael A. Powell (S'13) received the B.S. degree in mathematics from the United States Military Academy, West Point, NY, USA, in 2004, and the M.S.E. degree in biomedical engineering, in 2013, from The Johns Hopkins University, Baltimore, MD, USA, where he is currently working toward the Ph.D. degree.

He is an active duty U.S. Army officer. His work in the Neuroengineering and Biomedical Instrumentation Lab focuses on implementing pattern recognition control for myoelectric prostheses. He routinely works with amputees to develop and test new EMG decoding algorithms and prosthesis control strategies.



Rahul R. Kaliki (M'13) received the B.S. degree in biomedical engineering from the University of California-San Diego, La Jolla, CA, USA, in 2004, and the M.S. and Ph.D. degrees from the University of Southern California, Los Angeles, CA, USA, in 2009.

After finishing the Ph.D. degree, he joined Infinite Biomedical Technologies (IBT), Baltimore, MD, USA, as a Research Scientist and subsequently became the Chief Executive Officer in 2011. He is also a Visiting Scientist at The Johns Hopkins University, Baltimore, MD, USA. His team at IBT is focused on developing and commercializing products to improve the quality of life for people with limb amputations.



Nitish V. Thakor (F'94) is a Professor of Biomedical Engineering, Electrical and Computer Engineering, and Neurology at The Johns Hopkins University, and directs the Laboratory for Neuroengineering. He is also the Director of the Singapore Institute for Neurotechnology (SINAPSE) at the National University of Singapore. His technical expertise is in the field of neuroengineering, including neural diagnostic instrumentation, neural microsystems, neural signal processing, optical imaging of the nervous system, neural control of prostheses, and

brain-machine interfaces.

Dr. Thakor is a recipient of a Research Career Development Award from the National Institutes of Health and a Presidential Young Investigator Award from the National Science Foundation, and is a Fellow of the American Institute of Medical and Biological Engineering, Founding Fellow of the Biomedical Engineering Society, and Fellow of International Federation of Medical and Biological Engineering.