Individuals with Upper Limb Loss Require Minimal Training to Achieve Robust Motion Classification Using Sonomyography

Susannah Engdahl[1†], Ananya Dhawan[1†], Ahmed Bashatah[1], Guoqing Diao[2], Biswarup Mukherjee[1,3], Brian Monroe[4], Rahsaan Holley[5], Siddhartha Sikdar[1,3*]

[1] Department of Bioengineering, George Mason University, Fairfax, VA
[2] Department of Biostatistics and Bioinformatics, The George Washington University, Washington, DC
[3] Center for Adaptive Systems of Brain-Body Interactions, George Mason University, Fairfax, VA
[4] Hanger Clinic, Laurel, MD
[5] MedStar National Rehabilitation Hospital, Washington, DC

[†] Authors contributed equally
* Corresponding author

Correspondence to: ssikdar@gmu.edu

Abstract

Background: Although surface electromyography is commonly used as a sensing strategy for upper limb prostheses, it remains difficult to reliably decode the recorded signals for controlling multi-articulated hands. Sonomyography, or ultrasound-based sensing of muscle deformation, overcomes some of these issues and allows individuals with upper limb loss to reliably perform multiple motion patterns. The purposes of this study were to determine 1) the effect of training on classification performance with sonomyographic control and 2) the effect of training on the underlying muscle deformation patterns.

Methods: A series of motion pattern datasets were collected from five individuals with transradial limb loss. Each dataset contained five ultrasound images corresponding each of the following five motions: power grasp, wrist pronation, key grasp, tripod, point. Participants initially performed the motions for the datasets without receiving feedback on their performance (*baseline* phase), then with visual and verbal feedback (*feedback* phase), and finally again without feedback (*retention* phase). Cross-validation accuracy and metrics describing the consistency and separability of the muscle deformation patters were computed for each dataset. Changes in classification performance over the course of the study were assessed using linear mixed models. Associations between classification performance and the consistency and separability metrics were evaluated using Pearson correlations.

Results: The average cross-validation accuracy for each phase was 92% or greater and there was no significant change in cross-validation accuracy throughout training. Misclassifications of one motion as another did not persist systematically across datasets. Few of the correlations were significant, although many were moderate or greater in strength and showed a positive association between accuracy and improved consistency and separability metrics.

Conclusions: Participants were able to achieve high classification rates upon their initial exposure to sonomyography and training did not affect their performance. Thus, motion classification using sonomyography may be highly intuitive and is unlikely to require a structured training protocol to gain proficiency.

Keywords: sonomyography, prosthesis control, gesture recognition, motor learning, training

Background

Despite the enormous investment of resources in the development of new multi-articulated upper limb prosthetics, a large proportion of individuals with upper limb loss discontinue use of their prosthesis (1–3). Users often experience dissatisfaction with the function and control of the prosthesis (4,5), which may explain why a majority of non-users (88%) report the systems are “too difficult or tiring” to use (6). Although 74% of those who have abandoned their upper limb prosthesis state that they would reconsider their decision if improvements in functionality and usability were made (6), it remains technically challenging to robustly infer a user’s volitional motor intent for controlling a dexterous prosthesis. This could contribute to the lack of difference in user satisfaction between terminal devices having single or multiple degrees of freedom (7).

Surface electromyography (EMG) has historically been the predominant method for sensing muscle activation, but it is limited by poor amplitude resolution and low signal-to-noise ratio, especially with dry electrodes used in prosthesis sockets (8,9). There is also low specificity between muscles, especially in the forearm, due to cross-talk and co-activation (10–13). Consequently, multi-articulated hands tend to rely on direct control strategies for opening and closing the terminal device in which EMG signals are recorded from an agonist-antagonist muscle pair. Mode-switching is then used to toggle between different grasp patterns, requiring a special EMG trigger (e.g., co-contraction), physical gesture, or button press to initiate the switch. These functions are time-consuming and unintuitive, leading some users to report that they strongly dislike mode switching (14).

An alternative approach for controlling multi-articulated prosthetic hands relies on pattern recognition to decode user intent from EMG signal patterns. Pattern recognition has received considerable attention in the literature and there are currently three commercially-available systems in the United States with FDA clearance (COMPLETE CONTROL - Coapt; Sense - Infinite Biomedical Technologies; Myo Plus - Ottobock). Although pattern recognition algorithms enable successful real-time grasp classification (15–17) and allow for control of a prosthetic hand during real-world functional tasks (18–21), users and therapists both report that extended periods of training are typically necessary to achieve stable performance (22,23). It is important to note that training to use a prosthesis involves multiple stages—including learning about the underlying functioning of the control system and how to generate the requisite control signals (*conceptual training*), learning to control a physical prosthesis (*control training*), and learning to use the prosthesis for completing functional tasks (*functional training*) (24). In this paper, we will restrict our discussion to the earliest phase of training.

Although all training phases appear to be difficult for patients who are new to pattern recognition, learning to produce the control signals can be especially problematic. During this phase, the user must learn to produce a specialized set of EMG patterns that are sufficiently consistent and separable from each other to permit accurate classification. This can be difficult since people generally do not have experience modulating EMG signal amplitudes (25).

Individuals with limb loss may be further disadvantaged by motor cortex reorganization following amputation (26), as well as muscle atrophy due to disuse of the residual limb and/or increased reliance on the intact limb (27). Given these difficulties, it is unsurprising that first attempts to use pattern recognition are often error-prone. For example, initial classification accuracies for individuals with transradial limb loss have been reported to range from 46.37% (28) to 77.5% (29). Training over the course of multiple sessions or days appears to mitigate some of these errors for individuals with and without limb loss, regardless of whether feedback on their performance is provided (29–32). These improvements are credited to changes in the EMG signal patterns such that they become more consistent and/or separable, although the correlation between performance and EMG pattern characteristics is complex and not yet fully understood (32).

In order to overcome these problems with myoelectric control, some researchers are exploring the use of sonomyography (SMG), or ultrasound-based sensing of muscle deformations. This modality avoids many of the limitations of EMG because it can spatially resolve individual muscles with sub-millimeter precision, including those in deep-seated muscle compartments. As a result, cross-talk is effectively suppressed and the control signals derived from the detected muscle activity have a high signal-to-noise ratio. Prior studies have demonstrated clear potential for the use of SMG in controlling a prosthesis or other human-machine interface (33–40). Our own work has shown that SMG is capable of accurately classifying motor intent for able-bodied individuals (41,42) and individuals with upper limb loss (43,44) in both offline and real-time settings. Importantly, individuals with upper limb loss are able to achieve 96% classification accuracy for five grasps after only a few minutes of training (44).

Given the known challenges with prolonged training times required for EMG pattern recognition and the promising initial results with SMG, we wanted to systematically investigate whether it was feasible for individuals with upper limb loss to generate the requisite control signals and demonstrate proficiency with SMG motion classification in a single session. In particular, it is unclear whether training with SMG will lead to any improvements in classification performance over time. The relationship between performance and muscle deformation pattern characteristics (i.e., separability between patterns of different motions and consistency within patterns of the same motion) is also currently unknown. Therefore, the primary goal of this study was to determine the effect of training on SMG classification performance. We hypothesized that providing feedback during training would improve the classification accuracy and that this improvement would be retained if the feedback was removed. A secondary goal of this study was to determine the effect of training on muscle deformation patterns. We hypothesized that the patterns would change with training such that they correlate with the classification accuracy.

Methods

Subjects

We recruited five individuals with unilateral transradial limb amputation (Table 1). All individuals reported using myoelectric prostheses at the time of data collection. Some participants had experienced the use of SMG for one or two sessions while participating in other studies, but at least nine months had elapsed between their prior exposure and participation in the current study. All participants provided written informed consent prior to participating in this institutionally approved study.

[INSERT TABLE 1 ABOUT HERE]

Experimental protocol

Participants were instrumented with a clinical ultrasound system (Terason uSmart 3200T, Terason, Burlington, MA). A low-profile, high-frequency, linear 16HL7 ultrasound transducer was positioned on the volar aspect of the residual limb using a stretchable fabric cuff such that muscle deformations associated with all individual phantom finger movements were visually identifiable on the ultrasound images. Ultrasound image sequences were acquired and transferred to a PC in real-time using a USB-based video grabber (DVI2USB 3.0, Epiphan Systems, Inc.). The captured screen was then downscaled to 100 × 140 pixels to include only the relevant ultrasound image. The acquired image frames were processed in MATLAB (MathWorks, Natick, MA) using custom algorithms.

A series of datasets were collected from participants following the procedures outlined previously (44). For each dataset, participants performed repeated iterations of one motion from a set of five (power grasp, wrist pronation, key grasp, tripod, point). Starting from a resting position, they followed a cue and moved towards the end state of the desired motion over the course of one second, held the end state position for one second, moved back to rest over the course of one second, and remained at rest for one second. After they repeated this process five times in succession, we extracted the ultrasound image frames corresponding to the motion end state and rest. The extracted frames were averaged into a single image representing the motion end state or rest, which was then added to the dataset. This process was repeated until all five motions were included in the dataset. Once the dataset was completed, we performed leave-one-out cross-validation with a modified 1-nearest-neighbor classifier that used Pearson's correlation coefficient as a similarity measure, following an algorithm described in more detail previously (42). Our modified classifier averages the similarity measurements by class and selects the most similar class instead of selecting the most similar individual image.

The datasets were collected in three different phases. During the first phase (*baseline*), participants were not given any feedback about their performance in order to evaluate how well they could intuitively use the system. Three datasets were collected in this phase. During the second phase (*feedback*), participants were given visual and verbal feedback about their performance. Participants were told that the system was sensing their muscle deformation and were allowed to view the raw ultrasound images in real-time for context. They were also given a visual cue to help them monitor the extent of muscle deformation detected by the system. For each movement sequence, we calculated the Pearson correlation between the first ultrasound image frame (corresponding to a rest state) and the incoming image. The correlation value was inverted and graphically displayed in real-time such that peaks corresponded to high muscle deformation (i.e., dissimilarity from rest) and valleys corresponded to low muscle deformation (i.e., similarity to rest). This real-time display has been described elsewhere (44). After each dataset was created, participants were told the results of the cross-validation and were shown the associated confusion matrix to help them understand the source of any errors. They were also given suggestions on how they could alter their movements to try and improve the classification accuracy. Depending on the amount of time participants had available for data collection, either three or five datasets were collected during this phase. During the final phase (*retention*), three more datasets were collected in the absence of any feedback in order to evaluate whether any changes in performance with feedback were maintained.

All datasets were collected in succession on a single day without repositioning the ultrasound transducer, except for Am7. He had a longer testing session because he needed the motion performance speed to be slowed from one second to two seconds per cue. As a result, he required a break between collection of the third and fourth datasets and requested to have the transducer removed. Additionally, Am5 terminated the testing session early because of a scheduling conflict before datasets in the *retention* phase were collected. He returned three days later to complete a full testing session. Data for both sessions were retained for analysis in this case.

Data analysis

The primary outcome metric was cross-validation accuracy (Eq. 1), defined as the percent of data correctly classified during the leave-one out validation process for a given dataset *i*:

$$CA_i = 100 * \frac{P_{correct_i}}{P_{total_i}}, \tag{1}$$

where $P_{correct_i}$ is the correct number of predictions by the closest-class classifier and P_{total_i} is the total number of predictions (i.e., the total number of datapoints).

Cross-validation accuracy is a combined measurement of the user's ability to perform a motion and the classifier's ability to label individual motion performances. Since user performance and classifier performance are inherently linked in this metric, it is possible that a user's performance could change over time without affecting the cross-validation accuracy. For example, a user may perform the tripod grasp with very little variation for a given dataset, resulting in a high cross-validation accuracy. On the next dataset, they may perform the grasp with two different variations having slightly different levels of middle finger flexion. As long as the closest identified class for each of the variations is still tripod, the cross-validation accuracy would be unaffected. Therefore, in order to more appropriately understand the changes in user performance independent of the classifier performance, we represented the 100 x 140 pixel images in our dataset as points in 14,000-dimensional space such that each pixel in the image corresponds to an axis in the high dimensional space. We can then define point clusters in this high dimensional space such that each cluster is comprised of all the points in an associated motion class. We utilized metrics from the unsupervised learning literature to describe the characteristics of these clusters, and thus of the performances of each motion.

The clustering metrics used in our analysis include Caliński-Harabasz (CH) Index (45), the Silhouette Index (46), and the S_Dbw Index (47). The CH Index and Silhouette Index are both commonly used in the unsupervised learning literature, while the S_Dbw Index is less common but has been shown to be more robust (48). Each of these metrics is a combination of some measurement of cluster consistency and cluster separability. As such, we also discuss these constituent components (consistency and separability) as supplementary metrics to better understand inter-motion vs intra-motion behavior independently from each other. If a user's performance of a given motion becomes more consistent with other performances of the same motion, the points in that motion cluster move closer together and the consistency measurements improve. If a user's performance of a given motion becomes more distinct from the performances of another motion, the clusters themselves move further apart and the separability metrics improve. A more detailed explanation of the clustering metrics and their consistency and separability constituents is provided below.

<u>CH Index</u>

The CH Index is a positive unbounded measurement where higher values indicate more consistent and/or separable clusters. It has been shown to be robust when evaluating clusters that may have varying densities and may be comprised of subclusters themselves, but can be susceptible to errors when evaluating clusters with noise in the data or clusters with imbalance (48). The CH Index (Eq. 2) is defined as a ratio of the variance of the cluster centroids (CH-Separability) to the variance within each cluster (CH-Consistency):

$$CH\ Index = \frac{(N-k)}{(k-1)} \times \frac{\sum_i^k CHS_i}{\sum_i^k CHC_i}, \qquad (2)$$

where N is the number of datapoints, k is the number of clusters, CHS_i is the CH-Separability metric for a cluster C_i, and CHC_i is the CH-Consistency metric for C_i,. CH-Separability and CH-Consistency are defined as:

$$CHS_i = N_i \|\mu_i - \mu\|^2 \qquad (2.1)$$

$$CHC_i = \sum_{x \in C_i} \|x - \mu_i\|^2, \qquad (2.2)$$

where N_i is the number of points in C_i, μ_i is the cluster centroid of C_i (average of all points in C_i), μ is the data centroid or the average of all datapoints, and x is a datapoint in a given cluster, C_i.

Silhouette Index

The Silhouette Index is a bounded measurement between [-1 and 1] where higher values indicate more consistent and/or separable clusters. It has been shown to be robust when evaluating clusters with noise in the data, clusters that may have varying densities and clusters with imbalance, but can be susceptible to errors when evaluating clusters that may be comprised of subclusters themselves (48). The Silhouette Index (Eq. 3) is defined on a per point basis and summed across all N datapoints in a given dataset D:

$$Silhouette\ Index = \frac{1}{N}\sum_{x\in D}\frac{SS(x)-SC(x))}{\max(\ SS(x),\ \ SC(x))}, \quad (3)$$

where *SS(x)* represents the separability for a given datapoint *x* and *SC(x)* represents the consistency for a given datapoint *x*. *SS(x)* is defined as the minimum over clusters of the average distance between *x* and points in another cluster:

$$SS(x) = \min_{1\le i\le k}\left(C'_x + \frac{1}{N_i}\sum_{x'\in Ci}\|x' - x\|\right), where\ C'_x = \begin{cases}\infty, if\ x\in C_i\\ 0, otherwise\end{cases}, \quad (3.1)$$

where N_i is the number of points in C_i, *k* is the number of clusters, and C'_x ensures that the *SS* calculation only considers distances to neighboring cluster and ignore distances to points in the same cluster as *x*. *SC(x)* is defined as the average distance between *x* and points in its own cluster:

$$SC(x) = \sum_{i=1}^{k} C_x \times \frac{1}{N_i-1}\sum_{x'\in Ci}\|x' - x\|\ , where\ C_x = \begin{cases}1, if\ x\in C_i\\ 0, otherwise\end{cases}, \quad (3.2)$$

where N_i is the number of points in $C_{i,}$ *and k* is the number of clusters, and C_x is a selection variable to ensure the distances in *SC* are computed only to points in the same cluster as *x*.

Although the Silhouette Index is defined on a per point basis, we represent the separability and consistency constituents as averages across all clusters. Thus, the Silhouette-Separability score SS_i is defined as the average separability over all points in a given cluster C_i :

$$SS_i = \frac{1}{N_i}\sum_{x\in C_i} SS(x). \quad (3.3)$$

Similarly, the Silhouette-Consistency score SC_i is defined as the average consistency over all points in a given cluster C_i :

$$SC_i = \frac{1}{N_i}\sum_{x\in C_i} SC(x). \quad (3.4)$$

<u>S_Dbw Index</u>

The S_Dbw Index is a positive unbounded measurement where lower values indicate more consistent and/or separable clusters. It has been shown to be robust when evaluating clusters with noise in the data, clusters that may have varying densities, clusters that may be comprised of subclusters themselves, and clusters with differences in the number of points in each cluster (48). The S_Dbw Index (Eq. 4) is defined as the sum of a scatter measurement (S_Dbw-Consistency) and a density between clusters measurement (S-Dbw-Separability):

$$S\_Dbw\ Index = \frac{1}{k}\sum_{i=1}^{k}(SDC_i + SDS_i), \tag{4}$$

where k is the number of clusters, SDC_i is the S-Dbw-Consistency score for a given cluster C_i, and SDS_i is the S-Dbw-Separability score for a given cluster C_i. SDC_i is defined as the magnitude of the variance of C_i divided by the magnitude of variance of the whole dataset D:

$$SDC_i = \frac{\|\sigma^2(C_i)\|}{\|\sigma^2(D)\|}. \tag{4.1}$$

SDS_i is defined as the average of a ratio of densities $d(m, C_{ij})$:

$$SDS_i = \frac{1}{k-1}\sum_{j=1, j\neq i}^{k}\frac{d(m_{ij}, C_{ij})}{\max(\ d(\mu_i, C_{ij}),\ \ d(\mu_{j,}, C_{ij})\ )} \tag{4.2}$$

$$d(m, C_{ij}) = \sum_{x \in C_{ij}} f(x,m)\ , where\ f(x,m) = \begin{cases} 0, if\ \|x-m\| > \overline{\sigma}(D) \\ 1, otherwise \end{cases} \tag{4.2.1}$$

where C_{ij} is the union of clusters C_i and C_j, μ_i is the centroid of cluster C_i, μ_j is the centroid of cluster C_j, and m_{ij} is the midpoint between μ_i and μ_j. The density $d(m, C_{ij})$ is the number of points in C_{ij} that are within the neighborhood of a point m, where the neighborhood size $\overline{\sigma}(D)$ is defined as the average across all clusters in D of the magnitude of the standard deviation of each cluster. SDS_i values are high when there are more points around the midpoint of two clusters than there are around either one of the individual cluster centroids.

Statistical analysis

To determine the effect of training and feedback on classification performance, we fit the following linear mixed model:

$$Y_{ij} = \beta_0 + b_i + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \beta_3 X_{ij3} + \epsilon_{ij}, \quad (5)$$

where Y_{ij} is the overall cross-validation accuracy for the *i*th subject on the *j*th dataset. In this model, X_{ij1} is the normalized elapsed time since the collection of the first dataset, $X_{ij2} = 1$ if the *j*th dataset is in the *feedback* phase and $X_{ij2} = 0$ otherwise, $X_{ij3} = 1$ if the *j*th dataset is in the *retention* phase and $X_{ij3} = 0$ otherwise, ϵ_{ij} is the residual error, and b_i is a random intercept accounting for within-subject correlations among repeated measures. Both b_i and ϵ_{ij} are assumed to be normally distributed and independent. The *baseline* phase is treated as the reference level. To account for the small sample size and potential violation of the model assumptions, we used the permutation test (49) to assess the significance of the effects of training and feedback on the overall cross-validation accuracy ($\alpha = 0.05$).

To determine the effect of training and feedback on the muscle deformation patterns, we first calculated the average overall cross-validation accuracy and the average accuracy per grasp across subjects for each of the three phases. We then calculated the change in accuracy between the *baseline* and *feedback* phases, as well as the *baseline* and *retention* phases. Similarly, we calculated the change in the clustering metrics between phases. Finally, we calculated Pearson's correlation coefficients between the changes in these metrics and the changes in accuracy rates. To account for the small sample size, we used the permutation test (49) to test the null hypothesis that there was no correlation ($\alpha = 0.05$).

Results

Cross-validation accuracy exceeded 76% for all 57 datasets collected across all subjects and was at least 92% for 45 of the datasets. Furthermore, the average cross-validation accuracy for each phase was at least 92% (baseline: 94.4 ± 3.1%; feedback: 95.4 ± 3.6%; retention: 92.0 ± 7.1%; Figure 1). The entire data collection session was somewhat lengthy since at least nine datasets per participant were collected, with the exception of the first session for Am5. The average elapsed time for the five completed collections was 97 ± 47 minutes (Figure 2), although this value is elevated by the unusually long testing time for Am7 (190 minutes). The elapsed time was much more consistent across the remaining subjects (86 ± 8 minutes). Ultimately, the linear mixed model showed no significant effect of phase or elapsed time on overall cross-validation accuracy (Table 2).

[INSERT TABLE 2 ABOUT HERE]

Although overall cross-validation accuracy for all five grasps was generally high, inspection of the accuracy rates for individual grasps reveals that there was occasional misclassification. However, visual inspection of the misclassification rates across all collected datasets for individuals participant shows no obvious patterns over time (Figure 3, Additional File 1).

There were few significant correlations between changes in cross-validation accuracy and changes in the clustering metrics (Table 3, Table 4). The change in accuracy for the tripod grasp between the *baseline* and *feedback* phases was significantly correlated with change in the S_Dbw Index (r = -0.896, p = 0.045) and S_Dbw-Consistency (r = -0.963, p = 0.031), while the change in overall accuracy was significantly correlated with change in the CH Index (r = 0.981, p = 0.016). Between the *baseline* and *retention* phases, the change in accuracy for the point grasp was correlated with the change in Silhouette-Separability (r = 0.884, p = 0.044) and S_Dbw-Consistency (r = 0.947, p = 0.027).

[INSERT TABLE 3 ABOUT HERE]

[INSERT TABLE 4 ABOUT HERE]

Although no other correlations were statistically significant, many were moderate or greater in strength ($|r| > 0.5$). Furthermore, the direction of the correlations was generally consistent with our expectation that improvements in accuracy would relate to improvements in clustering behavior. There were a few statistically insignificant exceptions (Table 3, Table 4), but their relevance cannot be determined based on the current study due to the small sample size.

Discussion

The primary purpose of this study was to determine the effect of training on classification performance. Although we expected that performance would improve with provision of feedback during training, our results did not support this hypothesis. In fact, all subjects were able to achieve successful classification without any instruction right after they started the study (average accuracy in the *baseline* phase = 94.4 ± 3.1%), and the classification performance did not systematically change over time or with the provision of feedback during training. Most participants experienced small fluctuations in classification performance between datasets due to isolated misclassifications (Additional File 1). These reductions in performance are more likely caused by movement of the transducer, minor variations in muscle contraction patterns, or other transient issues that would not preclude individuals with upper limb loss from successfully using SMG for motion classification.

These findings stand in contrast to previous research exploring the effect of training on classification performance in pattern recognition systems. The training protocols used in these studies have been lengthy, involving practice over multiple sessions or days with (29,32) or without (30,31) the provision of external feedback in order to improve classification performance. The necessity of including individualized coaching as part of a training protocol for pattern recognition remains unclear (32), but extensive training is a time-consuming and costly undertaking in either case. However, we anticipate that the benefit of undergoing a structured training protocol for learning to generate consistent and separable SMG signals would be low. A reduced need for this initial training could enable patients to devote more resources towards functional training with a physical prosthesis, which may still require involvement from a therapist. Interestingly, preliminary work from our group suggests that the need for functional training may also be reduced for patients using SMG compared to other control strategies. Am3 was able to operate a sonomyographic prosthesis and complete a functional task immediately after donning it, despite receiving no specific instructions on how to approach the task. Although Am3 was an experienced user of a direct control myoelectric prosthesis, his performance with the sonomyographic prosthesis was visibly smoother and involved less compensatory movement (Additional File 3).

A reduced need for controls and/or functional training may ultimately help diminish barriers to prosthesis access in the United States, where few clinicians specialize in caring for people with upper limb loss or have experience with justifying a course of treatment to insurers (50). For these reasons, it is perhaps unsurprising that one survey found that 35% of those with unilateral upper limb loss received no training of any kind and only 22% received more than 10 hours of training from a prosthetist or therapist (51). Unfortunately, therapy is an essential component of the rehabilitation process and the receipt of training to use a first prosthesis has been associated with increased satisfaction (7). With a more intuitive control strategy enabled by SMG, there may be a potential for increased satisfaction without the need for extensive involvement from a therapist. Experiencing an early sense of accomplishment from successfully learning the control strategy may also motivate users to continue practicing with the prosthesis and could reduce the likelihood that they abandon prosthesis use.

Although our participants had fairly similar classification performances, it is worth highlighting individual participant performances in order to exemplify some advantages of SMG. In particular, Am5 was fully naïve to the use of SMG during his first data collection session but achieved perfect classification on the first dataset prior to receiving any feedback. He also maintained an average of 99% accuracy across all six datasets. Am7 was also fully naïve to the use of SMG, but had slightly poorer classification performance in comparison to the other participants and required nearly double the amount of time to create each dataset. Based on comments from Am7 and our observation of his SMG signals, it appears that he had a difficult time relaxing his muscles to a “resting” position in between the repeated grasps. Am7 had undergone amputation slightly over a year before participating in this study and was a highly inexperienced myoelectric prosthesis user, having only owned his prosthesis for one week. He had significant muscle atrophy in his residual limb as a result of this disuse, which may have contributed to his difficulties. While this could suggest that having general familiarity with prosthesis use may impact an individual’s ability to produce appropriate SMG signals, it actually seems that SMG motion classification can be easily learned even by those lacking prior prosthesis experience. Indeed, Am7 achieved an average cross-validation accuracy of 89% across nine datasets even as a novice prosthesis user.

Our finding that most participants were able to generate the requisite control signals on the very first dataset without being provided any instruction is indicative of the intuitive nature of SMG. Because SMG relies on sensing muscle deformations and these deformations are directly related to the proprioceptive afferents in muscle spindles, SMG control is highly congruent with the underlying proprioceptive sense in the residual limb musculature. Furthermore, the richness of the ultrasound features in high-dimensional space means that our algorithms can more easily distinguish the user's natural motion patterns. Users can therefore rely on proprioception and may not need other cues to monitor their performance. In fact, others have demonstrated that able-bodied individuals can successfully modulate the degree of muscle activation to a desired level even when relying only on the implicit feedback available through proprioception (52). Am8's results provide an interesting demonstration of this concept. Her average cross-validation accuracy was slightly worse during the *feedback* phase compared to the *baseline* and *retention* phases, suggesting that she performed best when relying on her own intuition rather than following explicit instructions. Although Am8 had one exposure to SMG five years prior to the current study, she regularly used a single degree-of-freedom myoelectric hand in her daily life and thus had minimal experience with gesture recognition to guide her performance.

The second purpose of this study was to determine the effect of training on muscle deformation patterns. We hypothesized that changes in the patterns would correlate with changes in classification accuracy over training, but we did not see this trend in our results (Additional File 2). Few of the pattern characteristics were significantly correlated with changes in classification accuracy—most notably, changes in the CH Index were correlated with changes in overall accuracy between the *baseline* and *feedback* phases. Classification was performed using a modified 1-nearest-neighbor classifier, in which motions are assigned to the nearest class within the high dimensional feature space. The CH Index, a ratio between the average inter-cluster and intra-cluster distances, is effectively a measure of distance between neighboring clusters. Thus, correlation between this distance-based metric and distance-based classification is to be expected.

Although most other correlations were statistically insignificant, there are a few interesting trends to note in the magnitude of the correlation coefficients. In particular, the magnitude tended to be slightly higher for the consistency metrics than the separability metrics, indicating that changes in accuracy could be related more closely to greater intra-motion consistency rather than greater inter-motion separability. This would mean that participants became more consistent in executing the grasps but did not actually change how they executed them relative to the other motion. This is well-aligned with the idea that the users may be able to rely on their innate proprioception when using SMG and that SMG is an intuitive control paradigm due to direct relationship with proprioception. In particular, the algorithm is able to decode the user’s intent without the user needing to adapt to the algorithm. Nonetheless, it must be re-emphasized that these correlations were nonsignificant and should not be overinterpreted.

Taken together, these findings suggest that participants were able to generate separable movements right away and were able to consistently repeat those movements without requiring external feedback. We believe this finding represents a significant advantage over pattern recognition control, which similarly requires that EMG signal patterns are distinct and repeatable. Unfortunately, people do not naturally have experience with modulating EMG patterns to meet these requirements (25), nor is it clear which EMG pattern characteristics are most relevant to classification performance (32). It is therefore difficult to know how to effectively train users on pattern recognition, possibly limiting user motivation or interest in adopting this technology. Delineating the relationship between signal pattern characteristics and classification performance appears to be less critical to an individual's success with SMG than EMG, as users seem capable of achieving successful classification without intervention.

There are several limitations to this study. First, the majority of our participants were not fully naïve to SMG prior to participation in this study. It is possible that this prior experience could have improved their performance on this protocol, but we believe this is unlikely given that a minimum of nine months had passed since their most recent exposure. They also returned to using a myoelectric prosthesis during the intervening time, which may have interfered with the retention of any motor learning or skills obtained in previous sessions.

Additionally, we tested the classification performance on a limited time scale while subjects remained stationary. It is well-known that EMG classification can degrade in response to changes in arm position, electrode shifting, sweating, muscle fatigue, or during changes in signal characteristics over time (53). Similar issues may occur with SMG classification, which would require users to retrain the classifier after some period of use. Even if these deteriorations in performance occur, it does not invalidate our current finding that users can initially achieve robust motion classification with minimal training.

Another limitation of this study was that we utilized a commercially-available ultrasound imaging system with an array transducer. For translation of SMG technology to practical prosthesis sockets, we anticipate utilizing single-element transducers with low power electronics. Our previous work has indicated that the classification accuracy with sparse sensing is not compromised (54). However, this result has yet to be validated in individuals with limb loss. We are currently developing fully-integrated prototype SMG systems and additional studies are planned in the future. It should also be noted that the reported classification accuracies were obtained using a 1-nearest neighbor classifier. We purposely utilized one of the simplest classifiers in an effort to decouple user performance from classifier performance. More sophisticated classifiers, such as linear discriminant analysis commonly used for EMG pattern recognition, are expected to provide improved classification accuracy for SMG data as well.

Finally, the sample size for this study was small. This may have reduced our ability to detect statistically significant results, especially for the correlations between classification performance and muscle deformation patterns. More of the correlations could prove to be significant if this protocol was replicated in a larger group of participants. Furthermore, we cannot fully distinguish between the effects of repetitive practice and provision of feedback on classification performance. There could have been an interaction between these factors, but we could not include an interaction term in the linear mixed model due to the small sample size. However, visual inspection of the results leads us to believe that an interaction term would not have been significant even if it was included. In addition to the small sample size, we only tested participants who had acquired limb loss. Individuals with congenital limb difference may have more difficulty learning to use SMG if they lack phantom hand sensations or have limited proprioceptive sense in their residual limb. Although our prior work showed that one individual with congenital limb difference achieved a cross-validation accuracy of 85% for four motions (44), we have not systematically tested whether this population can become achieve robust motion classification with minimal training.

Conclusion

This study provides the first systematic investigation of the effect of training on SMG classification performance. We showed that individuals with upper limb loss were able to immediately achieve accurate motion classification using SMG and that their performance did not change over time or with the provision of feedback. Additionally, changes in the muscle deformation patterns did not significantly correlate with changes in classification accuracy, which further substantiates the idea that users do not have to alter their strategy over time in order to achieve successful classification rates. Together, these findings suggest that structured training protocols are unlikely to facilitate substantial improvements in classification accuracy for users of SMG. This work represents an important step towards demonstrating the viability of controlling upper limb prostheses with SMG.

List of abbreviations

EMG = electromyography; SMG = sonomyography; CH = Caliński-Harabasz

Declarations

Ethics approval and consent to participate

All participants provided written informed consent prior to participation in the study. This study was approved by the Institutional Review Board at George Mason University (#492701).

Consent for publication

Details that could be used to identify individual participants are not presented in this work.

Availability of data and materials

The datasets used during the current study are available from the corresponding author on reasonable request.

Competing interests

Siddhartha Sikdar is an inventor on a US patent related to the use of ultrasound imaging to control artificial devices. The other authors declare that they have no competing interests.

Funding

This work was partially supported by the Office of the Assistant Secretary of Defense for Health Affairs, through the Peer Reviewed Orthopaedic Research Program under Award No. W81XWH-16-1-0722. This work was also partially supported by the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under Award No. U01EB027601. Opinions, interpretations, conclusions and recommendations are those of the authors and are not necessarily endorsed by the Department of Defense or National Institutes of Health.

Authors' contributions

AD, BMukherjee, and SS designed the experiment. AD, AB, and BMukherjee collected the data. AD analyzed the data. GD performed the statistical analysis. SE, AD, BMonroe, RH, and SS interpreted the data. SE drafted the manuscript. AD, AB, BMukherjee, BMonroe, RH, and SS revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to thank the participants for their involvement in this study.

References

1. Biddiss EA, Chau TT. Upper limb prosthesis use and abandonment: a survey of the last 25 years. Prosthet Orthot Int. 2007;31(3):236–57.
2. Østlie K, Lesjø IM, Franklin RJ, Garfelt B, Skjeldal OH, Magnus P. Prosthesis rejection in acquired major upper-limb amputees: a population-based survey. Disabil Rehabil Assist Technol. 2012;7(4):294–303.
3. McFarland LV, Hubbard Winkler SL, Heinemann AW, Jones M, Esquenazi A. Unilateral upper-limb loss: satisfaction and prosthetic-device use in veterans and servicemembers from Vietnam and OIF/OEF conflicts. J Rehabil Res Dev. 2010;47(4):299–316.
4. Kyberd PJ, Hill W. Survey of upper limb prosthesis users in Sweden, the United Kingdom and Canada. Prosthet Orthot Int. 2011;35(2):234–41.
5. Biddiss E, Beaton D, Chau T. Consumer design priorities for upper limb prosthetics. Disabil Rehabil Assist Technol. 2007;2(6):346–57.
6. Biddiss E, Chau T. Upper-limb prosthetics: critical factors in device abandonment. Am J Phys Med Rehabil. 2007;86(12):977–87.
7. Resnik L, Borgia M, Heinemann AW, Clark MA. Prosthesis satisfaction in a national sample of Veterans with upper limb amputation. Prosthet Orthot Int. 2020;44(2):81–91.
8. Clancy E, Morin EL, Merletti R. Sampling, noise-reduction and amplitude estimation issues in surface electromyography. J Electromyogr Kinesiol. 2002;12(1):1.

9. Graimann B, Dietl H. Introduction to upper limb prosthetics. In: Farina D, Jensen W, Akay M, editors. Introduction to Neural Engineering for Motor Rehabilitation. Hoboken, NJ: John Wiley & Sons, Inc.; 2013. p. 267–90.
10. Kong YK, Hallbeck MS, Jung MC. Crosstalk effect on surface electromyogram of the forearm flexors during a static grip task. J Electromyogr Kinesiol. 2010;20(6):1223–9.
11. van Duinen H, Gandevia SC, Taylor JL. Voluntary activation of the different compartments of the flexor digitorum profundus. J Neurophysiol. 2010;104(6):3213–21.
12. van Duinen H, Yu WS, Gandevia SC. Limited ability to extend the digits of the human hand independently with extensor digitorum. J Physiol. 2009;587(20):4799–810.
13. McIsaac TL, Fuglevand AJ. Motor-unit synchrony within and across compartments of the human flexor digitorum superficialis. J Neurophysiol. 2007;97(1):550–6.
14. Hargrove LJ, Scheme EJ, Englehart KB, Hudgins BS. Multiple binary classifications via linear discriminant analysis for improved controllability of a powered prosthesis. IEEE Trans Neural Syst Rehabil Eng. 2010;18(1):49–57.
15. Cipriani C, Antfolk C, Controzzi M, Lundborg G, Rosen B, Carrozza MC, et al. Online myoelectric control of a dexterous hand prosthesis by transradial amputees. IEEE Trans Neural Syst Rehabil Eng. 2011;19(3):260–70.
16. Jiang N, Rehbaum H, Vujaklija I, Graimann B, Farina D. Intuitive, online, simultaneous, and proportional myoelectric control over two degrees-of-freedom in upper limb amputees. IEEE Trans Neural Syst Rehabil Eng. 2014;22(3):501–10.
17. Simon AM, Hargrove LJ, Lock BA, Kuiken TA. Target Achievement Control Test: evaluating real-time myoelectric pattern-recognition control of multifunctional upper-limb prostheses. J Rehabil Res Dev. 2011;48(6):619–27.

18. Amsuess S, Vujaklija I, Goebel P, Roche AD, Graimann B, Aszmann OC, et al. Context-Dependent Upper Limb Prosthesis Control for Natural and Robust Use. IEEE Trans Neural Syst Rehabil Eng. 2016;24(7):744–53.

19. Kuiken TA, Miller LA, Turner K, Hargrove LJ. A Comparison of Pattern Recognition Control and Direct Control of a Multiple Degree-of-Freedom Transradial Prosthesis. IEEE J Transl Eng Health Med. 2016;4:1–8.

20. Resnik L, Huang HH, Winslow A, Crouch DL, Zhang F, Wolk N. Evaluation of EMG pattern recognition for upper limb prosthesis control: a case study in comparison with direct myoelectric control. J NeuroEngineering Rehabil. 2018;15(1):23.

21. Simon AM, Turner KL, Miller LA, Hargrove LJ, Kuiken TA. Pattern recognition and direct control home use of a multi-articulating hand prosthesis. In: 2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR). 2019. p. 386–91.

22. Franzke AW, Kristoffersen MB, Bongers RM, Murgia A, Pobatschnig B, Unglaube F, et al. Users' and therapists' perceptions of myoelectric multi-function upper limb prostheses with conventional and pattern recognition control. PLOS One. 2019;14(8):e0220899.

23. Resnik LJ, Acluche F, Klinger SL. User experience of controlling the DEKA Arm with EMG pattern recognition. PLOS One. 2018;13(9):e0203987.

24. Simon AM, Lock BA, Stubblefield KA. Patient Training for Functional Use of Pattern Recognition–Controlled Prostheses. J Prosthet Orthot. 2012;24(2).

25. Johnson RE, Kording KP, Hargrove LJ, Sensinger JW. EMG Versus Torque Control of Human–Machine Systems: Equalizing Control Signal Variability Does not Equalize Error or Uncertainty. IEEE Trans Neural Syst Rehabil Eng. 2017;25(6):660–7.

26. Di Pino G, Maravita A, Zollo L, Guglielmelli E, Di Lazzaro V. Augmentation-related brain plasticity. Front Syst Neurosci. 2014;8(109).

27. Chadwell A, Kenney L, Granat MH, Thies S, Head J, Galpin A, et al. Upper limb activity in myoelectric prosthesis users is biased towards the intact limb and appears unrelated to goal-directed task performance. Sci Rep. 2018;8(1):11084.

28. Atzori M, Gijsberts A, Castellini C, Caputo B, Hager A-GM, Elsig S, et al. Electromyography data for non-invasive naturally-controlled robotic hand prostheses. Sci Data. 2014;1(1):140053.

29. Powell M, Kaliki R, Thakor N. User Training for Pattern Recognition-Based Myoelectric Prostheses: Improving Phantom Limb Movement Consistency and Distinguishability. IEEE Trans Neural Syst Rehabil Eng. 2013;22(3):522–32.

30. Bunderson NE, Kuiken TA. Quantification of Feature Space Changes With Experience During Electromyogram Pattern Recognition Control. IEEE Trans Neural Syst Rehabil Eng. 2012;20(3):239–46.

31. He J, Zhang D, Jiang N, Sheng X, Farina D, Zhu X. User adaptation in long-term, open-loop myoelectric training: implications for EMG pattern recognition in prosthesis control. J Neural Eng. 2015;12(4):046005.

32. Kristoffersen MB, Franzke AW, Sluis CK van der, Murgia A, Bongers RM. The Effect of Feedback During Training Sessions on Learning Pattern-Recognition-Based Prosthesis Control. IEEE Trans Neural Syst Rehabil Eng. 2019;27(10):2087–96.

33. Zheng YP, Chan MMF, Shi J, Chen X, Huang QH. Sonomyography: Monitoring morphological changes of forearm muscles in actions with the feasibility for the control of powered prosthesis. Med Eng Phys. 2006;28(5):405–15.

34. Guo J, Zheng Y, Huang Q, Chen X. Dynamic monitoring of forearm muscles using one-dimensional sonomyography system. J Rehabil Res Dev. 2008;45(1):187.
35. Chen X, Zheng Y-P, Guo J-Y, Shi J. Sonomyography (SMG) Control for Powered Prosthetic Hand: A Study with Normal Subjects. Ultrasound Med Biol. 2010;36(7):1076–88.
36. Castellini C, Passig G, Zarka E. Using ultrasound images of the forearm to predict finger positions. IEEE Trans Neural Syst Rehabil Eng. 2012;20(6):788–97.
37. Shi J, Guo JY, Hu SX, Zheng YP. Recognition of finger flexion motion from ultrasound image: a feasibility study. Ultrasound Med Biol. 2012;38(10):1695–704.
38. Yang X, Sun X, Zhou D, Li Y, Liu H. Towards Wearable A-Mode Ultrasound Sensing for Real-Time Finger Motion Recognition. IEEE Trans Neural Syst Rehabil Eng. 2018;26(6):1199–208.
39. Yang X, Chen Z, Hettiarachchi N, Yan J, Liu H. A Wearable Ultrasound System for Sensing Muscular Morphological Deformations. IEEE Trans Syst Man Cybern Syst. 2019;1–10.
40. Yang X, Yan J, Chen Z, Ding H, Liu H. A Proportional Pattern Recognition Control Scheme for Wearable A-mode Ultrasound Sensing. IEEE Trans Ind Electron. 2020;67(1):800–8.
41. Sikdar S, Rangwala H, Eastlake E, Hunt I, Nelson A, Devanathan J, et al. Novel method for predicting dexterous individual finger movements by imaging muscle activity using a wearable ultrasonic system. IEEE Trans Neural Syst Rehabil Eng. 2014;22(1):69–76.

42. Akhlaghi N, Baker CA, Lahlou M, Zafar H, Murthy KG, Rangwala HS, et al. Real-time Classification of Hand Motions using Ultrasound Imaging of Forearm Muscles. IEEE Trans Biomed Eng. 2016;63(8):1687–98.
43. Baker CA, Akhlaghi N, Rangwala H, Kosecka J, Sikdar S. Real-time, ultrasound-based control of a virtual hand by a trans-radial amputee. In: Conference Proceedings of the IEEE Engineering in Medicine and Biology Society. 2016. p. 3219–22.
44. Dhawan AS, Mukherjee B, Patwardhan S, Akhlaghi N, Diao G, Levay G, et al. Proprioceptive Sonomyographic Control: A novel method for intuitive and proportional control of multiple degrees-of-freedom for individuals with upper extremity limb loss. Sci Rep. 2019;9(1):9499.
45. Caliński T, Harabasz J. A dendrite method for cluster analysis. Commun Stat. 1974;3(1):1–27.
46. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math. 1987;20:53–65.
47. Halkidi M, Vazirgiannis M. Clustering validity assessment: finding the optimal partitioning of a data set. In: Proceedings 2001 IEEE International Conference on Data Mining. 2001. p. 187–94.
48. Liu Y, Li Z, Xiong H, Gao X, Wu J. Understanding of internal clustering validation measures. In: 2010 IEEE International Conference on Data Mining. 2010. p. 911–6.
49. Higgins J. An Introduction to Modern Nonparametric Statistics. Pacific Grove, CA: Brooks/Cole; 2004.

50. National Academies of Sciences, Engineering, and Medicine. The promise of assistive technology to enhance activity and work participation. Jette AM, Spicer CM, Flaubert JL, editors. Washington, DC: National Academies Press; 2017.

51. Kestner S. Defining the Relationship between Prosthetic Wrist Function and Its Use in Performing Work Tasks and Activities of Daily Living. J Prosthet Orthot. 2006;18(3):80–6.

52. Markovic M, Schweisfurth MA, Engels LF, Farina D, Dosen S. Myocontrol is closed-loop control: incidental feedback is sufficient for scaling the prosthesis force in routine grasping. J NeuroEngineering Rehabil. 2018;15(1):81.

53. Kyranou I, Vijayakumar S, Erden MS. Causes of Performance Degradation in Non-invasive Electromyographic Pattern Recognition in Upper Limb Prostheses. Front Neurorobotics. 2018;12:58.

54. Akhlaghi N, Dhawan A, Khan A, Mukherjee B, Diao G, Truong C, et al. Sparsity Analysis of a Sonomyographic Muscle-Computer Interface. IEEE Trans Biomed Eng. 2019;67(3):688–96.

55. Hinterreiter A, Ruch P, Stitz H, Ennemoser M, Bernard J, Strobelt H, et al. ConfusionFlow: A model-agnostic visualization for temporal analysis of classifier confusion. IEEE Trans Vis Comput Graph. 2020;

Figure Captions

Figure 1 Average between-subjects (grey bars) and within-subject (colored bars) cross-validation accuracy for each phase. Error bars represent standard deviation.

Figure 2 Cross-validation accuracy as function of elapsed training time for individual participants. The three sections of each plot correspond to the *baseline*, *feedback*, and *retention* phases. The break between the third and fourth datasets for Am7 indicates that the transducer was removed and repositioned.

Figure 3 Classification performance across individual datasets for Am3 (top) and Am7 (bottom). The confusion matrices have been adapted to represent the temporal evolution of classification performance across all datasets (55). The squares in each confusion matrix are divided into *n* columns representing the *n* collected datasets for that subject. Thus, confusion between grasps for individual datasets is illustrated by the individual columns. For example, power grasp was identified correctly for four out of five motion instances during Am7's first dataset (yellow bar) and was confused with point on one instance (maroon bar).

Tables

Table 1. Participant characteristics.

ID	Sex	Age	Years since amputation	Prior prosthesis experience*	Affected limb	Dates of prior data collections using SMG	Date of data collection for current study
Am1	M	56	46	DC (experienced)	Right	Nov. 29, 2017	Oct. 10, 2018
Am3	M	68	50	DC (experienced), PR (novice)	Left	Oct. 24, 2017 Dec. 11, 2017	Sept. 11, 2018
Am5	M	38	1.5	DC (novice), PR (novice)	Right	n/a	Feb. 23, 2018† Feb. 26, 2018
Am7	M	70	1	DC (novice)	Left	n/a	Jan. 24, 2020
Am8	F	74	12	DC (experienced)	Right	Aug. 6, 2015	Feb. 4, 2020

* novice indicates less than two years of experience
† session terminated early
DC = direct control; PR = pattern recognition

Table 2. Results from the linear mixed model.

Parameter	Estimate	Standard Error	t value	Pr(>t)*
Normalized elapsed time (β_1)	3.1377	5.4201	.5789	0.322
Feedback phase (β_2)	-0.6185	3.0983	-0.1996	0.707
Retention phase (β_3)	-4.9912	4.7435	-1.0522	0.872

*one-sided p-values based on 1000 permutations

Table 3. Pearson correlation coefficients between changes in accuracy rates from *baseline* to *feedback* and changes in the separability and consistency metrics from *baseline* to *feedback*.

			Accuracy (key)	Accuracy (point)	Accuracy (power)	Accuracy (tripod)	Accuracy (wrist)	Accuracy (overall)
		S_Dbw Index	0.483†	-0.259	-0.449	-0.896*	-0.457	-0.563
		CH Index	0.099	0.142	0.692	0.262	-0.156†	0.981*
		Silhouette Index	-0.123†	0.327	0.509	0.625	0.185	0.849
Separability	**CH - Separability**	**Key grasp**	0.652	-	-	-	-	-
		Point	-	0.623	-	-	-	-
		Power grasp	-	-	0.219	-	-	-
		Tripod	-	-	-	0.593	-	-
		Wrist pronation	-	-	-	-	0.432	-
	Silhouette - Separability	**Key grasp**	0.031	-	-	-	-	-
		Point	-	0.19	-	-	-	-
		Power grasp	-	-	-0.204†	-	-	-
		Tripod	-	-	-	0.605	-	-
		Wrist pronation	-	-	-	-	0.55	-
	S_Dbw - Separability	**Key grasp**	0.531†	-	-	-	-	-
		Point	-	0.453†	-	-	-	-
		Power grasp	-	-	-0.33	-	-	-
		Tripod	-	-	-	-0.793	-	-
		Wrist pronation	-	-	-	-	-1	-
Consistency	**CH - Consistency**	**Key grasp**	-0.807	-	-	-	-	-
		Point	-	-0.457	-	-	-	-
		Power grasp	-	-	-0.791	-	-	-
		Tripod	-	-	-	-0.671	-	-
		Wrist pronation	-	-	-	-	-0.405	-
	Silhouette - Consistency	**Key grasp**	-0.798	-	-	-	-	-
		Point	-	-0.386	-	-	-	-
		Power grasp	-	-	-0.78	-	-	-
		Tripod	-	-	-	-0.741	-	-
		Wrist pronation	-	-	-	-	-0.471	-
	S_Dbw - Consistency	**Key grasp**	-0.909	-	-	-	-	-
		Point	-	-0.563	-	-	-	-
		Power grasp	-	-	-0.699	-	-	-
		Tripod	-	-	-	-0.963*	-	-
		Wrist pronation	-	-	-	-	-0.968	-

* $p < 0.05$ (based on 1000 permutations)

† unexpected direction of correlation

Table 4. Pearson correlation coefficients between changes in accuracy rates from *baseline* to *retention* and changes in the separability and consistency metrics from *baseline* to *retention*.

			Accuracy (key)	Accuracy (point)	Accuracy (power)	Accuracy (tripod)	Accuracy (wrist)	Accuracy (overall)
		S_Dbw Index	0.429[†]	-0.688	-0.598	-0.213	-0.368	-0.511
		CH Index	-0.145[†]	0.403	0.673	0.256	0.224	0.542
		Silhouette Index	-0.031[†]	0.667	0.711	0.392	-0.01[†]	0.775
Separability	**CH - Separability**	**Key grasp**	0.494	-	-	-	-	-
		Point	-	0.877	-	-	-	-
		Power grasp	-	-	0.352	-	-	-
		Tripod	-	-	-	-0.173[†]	-	-
		Wrist pronation	-	-	-	-	0.675	-
	Silhouette - Separability	**Key grasp**	0.11	-	-	-	-	-
		Point	-	0.884*	-	-	-	-
		Power grasp	-	-	0.096	-	-	-
		Tripod	-	-	-	-0.072[†]	-	-
		Wrist pronation	-	-	-	-	0.694	-
	S_Dbw - Separability	**Key grasp**	0.21[†]	-	-	-	-	-
		Point	-	-0.896	-	-	-	-
		Power grasp	-	-	-0.423	-	-	-
		Tripod	-	-	-	-0.123	-	-
		Wrist pronation	-	-	-	-	-1	-
Consistency	**CH - Consistency**	**Key grasp**	-0.621	-	-	-	-	-
		Point	-	0.585[†]	-	-	-	-
		Power grasp	-	-	-0.725	-	-	-
		Tripod	-	-	-	-0.148	-	-
		Wrist pronation	-	-	-	-	-0.602	-
	Silhouette - Consistency	**Key grasp**	-0.688	-	-	-	-	-
		Point	-	0.352[†]	-	-	-	-
		Power grasp	-	-	-0.751	-	-	-
		Tripod	-	-	-	-0.204	-	-
		Wrist pronation	-	-	-	-	-0.67	-
	S_Dbw - Consistency	**Key grasp**	-0.621	-	-	-	-	-
		Point	-	-0.947*	-	-	-	-
		Power grasp	-	-	-0.873	-	-	-
		Tripod	-	-	-	-0.536	-	-
		Wrist pronation	-	-	-	-	-0.879	-

* $p < 0.05$ (based on 1000 permutations)

[†] unexpected direction of correlation

Additional Files

Additional File 1 **Classification performance across individual datasets.** The squares in each confusion matrix are divided into *n* columns representing the *n* collected datasets for that subject.

Additional File 2 **Clustering metrics as function of elapsed training time for individual participants.** The S_Dbw Index is negated to facilitate comparison with the other clustering metrics. Increasing values correspond to increasing separability and/or consistency. The three sections of each plot correspond to the *baseline*, *feedback*, and *retention* phases.

Additional File 3 **Box and Blocks Test for Am3.** Am3 performed the test with his clinically-prescribed myoelectric prosthesis and an experimental sonomyographic prosthesis.