

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

Susannah Engdahl

George Mason University <https://orcid.org/0000-0001-7341-4624>

Ananya Dhawan

George Mason University

Ahmed Bashatah

George Mason University

Guoqing Diao

George Washington University

Biswarup Mukherjee

George Mason University

Brian Monroe

Hanger Clinic

Rahsaan Holley

MedStar National Rehabilitation Hospital

Siddhartha Sikdar (✉ ssikdar@gmu.edu)

George Mason University <https://orcid.org/0000-0002-6426-2320>

Research

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

Posted Date: September 9th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-65415/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

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

Susannah Engdahl^{1†}, Ananya Dhawan^{1†}, Ahmed Bashatah¹, Guoqing Diao², Biswarup Mukherjee^{1,3}, Brian Monroe⁴, Rahsaan Holley⁵, 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

1 **Abstract**

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

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

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

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

28

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

30

31 **Background**

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

42 Surface electromyography (EMG) has historically been the predominant method for
43 sensing muscle activation, but it is limited by poor amplitude resolution and low signal-to-noise
44 ratio, especially with dry electrodes used in prosthesis sockets (8,9). There is also low specificity
45 between muscles, especially in the forearm, due to cross-talk and co-activation (10–13).
46 Consequently, multi-articulated hands tend to rely on direct control strategies for opening and

47 closing the terminal device in which EMG signals are recorded from an agonist-antagonist
48 muscle pair. Mode-switching is then used to toggle between different grasp patterns, requiring a
49 special EMG trigger (e.g., co-contraction), physical gesture, or button press to initiate the switch.
50 These functions are time-consuming and unintuitive, leading some users to report that they
51 strongly dislike mode switching (14).

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

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

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

81 In order to overcome these problems with myoelectric control, some researchers are
82 exploring the use of sonomyography (SMG), or ultrasound-based sensing of muscle
83 deformations. This modality avoids many of the limitations of EMG because it can spatially
84 resolve individual muscles with sub-millimeter precision, including those in deep-seated muscle
85 compartments. As a result, cross-talk is effectively suppressed and the control signals derived
86 from the detected muscle activity have a high signal-to-noise ratio. Prior studies have
87 demonstrated clear potential for the use of SMG in controlling a prosthesis or other human-
88 machine interface (33–40). Our own work has shown that SMG is capable of accurately
89 classifying motor intent for able-bodied individuals (41,42) and individuals with upper limb loss
90 (43,44) in both offline and real-time settings. Importantly, individuals with upper limb loss are
91 able to achieve 96% classification accuracy for five grasps after only a few minutes of training
92 (44).

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

107

108 **Methods**

109 *Subjects*

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

116 [INSERT TABLE 1 ABOUT HERE]

117

118 *Experimental protocol*

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

128 A series of datasets were collected from participants following the procedures outlined
129 previously (44). For each dataset, participants performed repeated iterations of one motion from
130 a set of five (power grasp, wrist pronation, key grasp, tripod, point). Starting from a resting
131 position, they followed a cue and moved towards the end state of the desired motion over the
132 course of one second, held the end state position for one second, moved back to rest over the
133 course of one second, and remained at rest for one second. After they repeated this process five
134 times in succession, we extracted the ultrasound image frames corresponding to the motion end
135 state and rest. The extracted frames were averaged into a single image representing the motion
136 end state or rest, which was then added to the dataset. This process was repeated until all five
137 motions were included in the dataset. Once the dataset was completed, we performed leave-one-
138 out cross-validation with a modified 1-nearest-neighbor classifier that used Pearson's correlation

139 coefficient as a similarity measure, following an algorithm described in more detail previously
140 (42). Our modified classifier averages the similarity measurements by class and selects the most
141 similar class instead of selecting the most similar individual image.

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

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

169

170 *Data analysis*

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

$$CA_i = 100 * \frac{P_{correct_i}}{P_{total_i}}, \quad (1)$$

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

175 Cross-validation accuracy is a combined measurement of the user's ability to perform a
176 motion and the classifier's ability to label individual motion performances. Since user
177 performance and classifier performance are inherently linked in this metric, it is possible that a
178 user's performance could change over time without affecting the cross-validation accuracy. For
179 example, a user may perform the tripod grasp with very little variation for a given dataset,
180 resulting in a high cross-validation accuracy. On the next dataset, they may perform the grasp
181 with two different variations having slightly different levels of middle finger flexion. As long as
182 the closest identified class for each of the variations is still tripod, the cross-validation accuracy

183 would be unaffected. Therefore, in order to more appropriately understand the changes in user
184 performance independent of the classifier performance, we represented the 100 x 140 pixel
185 images in our dataset as points in 14,000-dimensional space such that each pixel in the image
186 corresponds to an axis in the high dimensional space. We can then define point clusters in this
187 high dimensional space such that each cluster is comprised of all the points in an associated
188 motion class. We utilized metrics from the unsupervised learning literature to describe the
189 characteristics of these clusters, and thus of the performances of each motion.

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

203

204 CH Index

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

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

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

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

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

214 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),
 215 μ is the data centroid or the average of all datapoints, and x is a datapoint in a given cluster, C_i .

216

217 Silhouette Index

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

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

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

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

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

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

231 where N_i is the number of points in C_i , and k is the number of clusters, and C_x is a selection
 232 variable to ensure the distances in SC are computed only to points in the same cluster as x .

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

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

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

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

238

239 S_Dbw Index

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

$$S_Dbw\ Index = \frac{1}{k} \sum_{i=1}^k (SDC_i + SDS_i), \quad (4)$$

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

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

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

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

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

250 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
 251 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
 252 points in C_{ij} that are within the neighborhood of a point m , where the neighborhood size $\bar{\sigma}(D)$ is
 253 defined as the average across all clusters in D of the magnitude of the standard deviation of each
 254 cluster. SDS_i values are high when there are more points around the midpoint of two clusters than
 255 there are around either one of the individual cluster centroids.

256

257 **Statistical analysis**

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

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

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

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

277

278 **Results**

279 Cross-validation accuracy exceeded 76% for all 57 datasets collected across all subjects
280 and was at least 92% for 45 of the datasets. Furthermore, the average cross-validation accuracy
281 for each phase was at least 92% (baseline: $94.4 \pm 3.1\%$; feedback: $95.4 \pm 3.6\%$; retention: $92.0 \pm$
282 7.1% ; Figure 1). The entire data collection session was somewhat lengthy since at least nine
283 datasets per participant were collected, with the exception of the first session for Am5. The
284 average elapsed time for the five completed collections was 97 ± 47 minutes (Figure 2), although
285 this value is elevated by the unusually long testing time for Am7 (190 minutes). The elapsed
286 time was much more consistent across the remaining subjects (86 ± 8 minutes). Ultimately, the
287 linear mixed model showed no significant effect of phase or elapsed time on overall cross-
288 validation accuracy (Table 2).

289 [INSERT TABLE 2 ABOUT HERE]

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

295 There were few significant correlations between changes in cross-validation accuracy and
296 changes in the clustering metrics (Table 3, Table 4). The change in accuracy for the tripod grasp
297 between the *baseline* and *feedback* phases was significantly correlated with change in the S_Dbw
298 Index ($r = -0.896$, $p = 0.045$) and S_Dbw-Consistency ($r = -0.963$, $p = 0.031$), while the change
299 in overall accuracy was significantly correlated with change in the CH Index ($r = 0.981$, $p =$
300 0.016). Between the *baseline* and *retention* phases, the change in accuracy for the point grasp

301 was correlated with the change in Silhouette-Separability ($r = 0.884$, $p = 0.044$) and S_Dbw-
302 Consistency ($r = 0.947$, $p = 0.027$).

303 [INSERT TABLE 3 ABOUT HERE]

304 [INSERT TABLE 4 ABOUT HERE]

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

310

311 **Discussion**

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

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

340 A reduced need for controls and/or functional training may ultimately help diminish
341 barriers to prosthesis access in the United States, where few clinicians specialize in caring for
342 people with upper limb loss or have experience with justifying a course of treatment to insurers
343 (50). For these reasons, it is perhaps unsurprising that one survey found that 35% of those with
344 unilateral upper limb loss received no training of any kind and only 22% received more than 10
345 hours of training from a prosthetist or therapist (51). Unfortunately, therapy is an essential

346 component of the rehabilitation process and the receipt of training to use a first prosthesis has
347 been associated with increased satisfaction (7). With a more intuitive control strategy enabled by
348 SMG, there may be a potential for increased satisfaction without the need for extensive
349 involvement from a therapist. Experiencing an early sense of accomplishment from successfully
350 learning the control strategy may also motivate users to continue practicing with the prosthesis
351 and could reduce the likelihood that they abandon prosthesis use.

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

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

386 The second purpose of this study was to determine the effect of training on muscle
387 deformation patterns. We hypothesized that changes in the patterns would correlate with changes
388 in classification accuracy over training, but we did not see this trend in our results (Additional
389 File 2). Few of the pattern characteristics were significantly correlated with changes in
390 classification accuracy—most notably, changes in the CH Index were correlated with changes in
391 overall accuracy between the *baseline* and *feedback* phases. Classification was performed using a

392 modified 1-nearest-neighbor classifier, in which motions are assigned to the nearest class within
393 the high dimensional feature space. The CH Index, a ratio between the average inter-cluster and
394 intra-cluster distances, is effectively a measure of distance between neighboring clusters. Thus,
395 correlation between this distance-based metric and distance-based classification is to be
396 expected.

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

408 Taken together, these findings suggest that participants were able to generate separable
409 movements right away and were able to consistently repeat those movements without requiring
410 external feedback. We believe this finding represents a significant advantage over pattern
411 recognition control, which similarly requires that EMG signal patterns are distinct and
412 repeatable. Unfortunately, people do not naturally have experience with modulating EMG
413 patterns to meet these requirements (25), nor is it clear which EMG pattern characteristics are
414 most relevant to classification performance (32). It is therefore difficult to know how to

415 effectively train users on pattern recognition, possibly limiting user motivation or interest in
416 adopting this technology. Delineating the relationship between signal pattern characteristics and
417 classification performance appears to be less critical to an individual's success with SMG than
418 EMG, as users seem capable of achieving successful classification without intervention.

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

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

432 Another limitation of this study was that we utilized a commercially-available ultrasound
433 imaging system with an array transducer. For translation of SMG technology to practical
434 prosthesis sockets, we anticipate utilizing single-element transducers with low power electronics.
435 Our previous work has indicated that the classification accuracy with sparse sensing is not
436 compromised (54). However, this result has yet to be validated in individuals with limb loss. We
437 are currently developing fully-integrated prototype SMG systems and additional studies are

438 planned in the future. It should also be noted that the reported classification accuracies were
439 obtained using a 1-nearest neighbor classifier. We purposely utilized one of the simplest
440 classifiers in an effort to decouple user performance from classifier performance. More
441 sophisticated classifiers, such as linear discriminant analysis commonly used for EMG pattern
442 recognition, are expected to provide improved classification accuracy for SMG data as well.

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

458

459 **Conclusion**

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

470

471 **List of abbreviations**

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

473

474 **Declarations**

475 *Ethics approval and consent to participate*

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

478

479 *Consent for publication*

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

481

482 *Availability of data and materials*

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

485

486 ***Competing interests***

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

489

490 ***Funding***

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

498

499 ***Authors' contributions***

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

504

505 ***Acknowledgements***

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

507

508 **References**

509 1. Biddiss EA, Chau TT. Upper limb prosthesis use and abandonment: a survey of the last 25
510 years. *Prosthet Orthot Int.* 2007;31(3):236–57.

511 2. Østlie K, Lesjø IM, Franklin RJ, Garfelt B, Skjeldal OH, Magnus P. Prosthesis rejection in
512 acquired major upper-limb amputees: a population-based survey. *Disabil Rehabil Assist*
513 *Technol.* 2012;7(4):294–303.

514 3. McFarland LV, Hubbard Winkler SL, Heinemann AW, Jones M, Esquenazi A. Unilateral
515 upper-limb loss: satisfaction and prosthetic-device use in veterans and servicemembers
516 from Vietnam and OIF/OEF conflicts. *J Rehabil Res Dev.* 2010;47(4):299–316.

517 4. Kyberd PJ, Hill W. Survey of upper limb prosthesis users in Sweden, the United Kingdom
518 and Canada. *Prosthet Orthot Int.* 2011;35(2):234–41.

519 5. Biddiss E, Beaton D, Chau T. Consumer design priorities for upper limb prosthetics.
520 *Disabil Rehabil Assist Technol.* 2007;2(6):346–57.

521 6. Biddiss E, Chau T. Upper-limb prosthetics: critical factors in device abandonment. *Am J*
522 *Phys Med Rehabil.* 2007;86(12):977–87.

523 7. Resnik L, Borgia M, Heinemann AW, Clark MA. Prosthesis satisfaction in a national
524 sample of Veterans with upper limb amputation. *Prosthet Orthot Int.* 2020;44(2):81–91.

525 8. Clancy E, Morin EL, Merletti R. Sampling, noise-reduction and amplitude estimation issues
526 in surface electromyography. *J Electromyogr Kinesiol.* 2002;12(1):1.

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

- 550 18. Amsuess S, Vujaklija I, Goebel P, Roche AD, Graimann B, Aszmann OC, et al. Context-
551 Dependent Upper Limb Prosthesis Control for Natural and Robust Use. *IEEE Trans Neural*
552 *Syst Rehabil Eng.* 2016;24(7):744–53.
- 553 19. Kuiken TA, Miller LA, Turner K, Hargrove LJ. A Comparison of Pattern Recognition
554 Control and Direct Control of a Multiple Degree-of-Freedom Transradial Prosthesis. *IEEE J*
555 *Transl Eng Health Med.* 2016;4:1–8.
- 556 20. Resnik L, Huang HH, Winslow A, Crouch DL, Zhang F, Wolk N. Evaluation of EMG
557 pattern recognition for upper limb prosthesis control: a case study in comparison with direct
558 myoelectric control. *J NeuroEngineering Rehabil.* 2018;15(1):23.
- 559 21. Simon AM, Turner KL, Miller LA, Hargrove LJ, Kuiken TA. Pattern recognition and direct
560 control home use of a multi-articulating hand prosthesis. In: 2019 IEEE 16th International
561 Conference on Rehabilitation Robotics (ICORR). 2019. p. 386–91.
- 562 22. Franzke AW, Kristoffersen MB, Bongers RM, Murgia A, Pobatschnig B, Unglaube F, et al.
563 Users’ and therapists’ perceptions of myoelectric multi-function upper limb prostheses with
564 conventional and pattern recognition control. *PLOS One.* 2019;14(8):e0220899.
- 565 23. Resnik LJ, Acluche F, Klinger SL. User experience of controlling the DEKA Arm with
566 EMG pattern recognition. *PLOS One.* 2018;13(9):e0203987.
- 567 24. Simon AM, Lock BA, Stubblefield KA. Patient Training for Functional Use of Pattern
568 Recognition–Controlled Prostheses. *J Prosthet Orthot.* 2012;24(2).
- 569 25. Johnson RE, Kording KP, Hargrove LJ, Sensinger JW. EMG Versus Torque Control of
570 Human–Machine Systems: Equalizing Control Signal Variability Does not Equalize Error
571 or Uncertainty. *IEEE Trans Neural Syst Rehabil Eng.* 2017;25(6):660–7.

- 572 26. Di Pino G, Maravita A, Zollo L, Guglielmelli E, Di Lazzaro V. Augmentation-related brain
573 plasticity. *Front Syst Neurosci.* 2014;8(109).
- 574 27. Chadwell A, Kenney L, Granat MH, Thies S, Head J, Galpin A, et al. Upper limb activity in
575 myoelectric prosthesis users is biased towards the intact limb and appears unrelated to goal-
576 directed task performance. *Sci Rep.* 2018;8(1):11084.
- 577 28. Atzori M, Gijssberts A, Castellini C, Caputo B, Hager A-GM, Elsig S, et al.
578 Electromyography data for non-invasive naturally-controlled robotic hand prostheses. *Sci*
579 *Data.* 2014;1(1):140053.
- 580 29. Powell M, Kaliki R, Thakor N. User Training for Pattern Recognition-Based Myoelectric
581 Prostheses: Improving Phantom Limb Movement Consistency and Distinguishability. *IEEE*
582 *Trans Neural Syst Rehabil Eng.* 2013;22(3):522–32.
- 583 30. Bunderson NE, Kuiken TA. Quantification of Feature Space Changes With Experience
584 During Electromyogram Pattern Recognition Control. *IEEE Trans Neural Syst Rehabil*
585 *Eng.* 2012;20(3):239–46.
- 586 31. He J, Zhang D, Jiang N, Sheng X, Farina D, Zhu X. User adaptation in long-term, open-
587 loop myoelectric training: implications for EMG pattern recognition in prosthesis control. *J*
588 *Neural Eng.* 2015;12(4):046005.
- 589 32. Kristoffersen MB, Franzke AW, Sluis CK van der, Murgia A, Bongers RM. The Effect of
590 Feedback During Training Sessions on Learning Pattern-Recognition-Based Prosthesis
591 Control. *IEEE Trans Neural Syst Rehabil Eng.* 2019;27(10):2087–96.
- 592 33. Zheng YP, Chan MMF, Shi J, Chen X, Huang QH. Sonomyography: Monitoring
593 morphological changes of forearm muscles in actions with the feasibility for the control of
594 powered prosthesis. *Med Eng Phys.* 2006;28(5):405–15.

- 595 34. Guo J, Zheng Y, Huang Q, Chen X. Dynamic monitoring of forearm muscles using one-
596 dimensional sonomyography system. *J Rehabil Res Dev.* 2008;45(1):187.
- 597 35. Chen X, Zheng Y-P, Guo J-Y, Shi J. Sonomyography (SMG) Control for Powered
598 Prosthetic Hand: A Study with Normal Subjects. *Ultrasound Med Biol.* 2010;36(7):1076–
599 88.
- 600 36. Castellini C, Passig G, Zarka E. Using ultrasound images of the forearm to predict finger
601 positions. *IEEE Trans Neural Syst Rehabil Eng.* 2012;20(6):788–97.
- 602 37. Shi J, Guo JY, Hu SX, Zheng YP. Recognition of finger flexion motion from ultrasound
603 image: a feasibility study. *Ultrasound Med Biol.* 2012;38(10):1695–704.
- 604 38. Yang X, Sun X, Zhou D, Li Y, Liu H. Towards Wearable A-Mode Ultrasound Sensing for
605 Real-Time Finger Motion Recognition. *IEEE Trans Neural Syst Rehabil Eng.*
606 2018;26(6):1199–208.
- 607 39. Yang X, Chen Z, Hettiarachchi N, Yan J, Liu H. A Wearable Ultrasound System for
608 Sensing Muscular Morphological Deformations. *IEEE Trans Syst Man Cybern Syst.*
609 2019;1–10.
- 610 40. Yang X, Yan J, Chen Z, Ding H, Liu H. A Proportional Pattern Recognition Control
611 Scheme for Wearable A-mode Ultrasound Sensing. *IEEE Trans Ind Electron.*
612 2020;67(1):800–8.
- 613 41. Sikdar S, Rangwala H, Eastlake E, Hunt I, Nelson A, Devanathan J, et al. Novel method for
614 predicting dexterous individual finger movements by imaging muscle activity using a
615 wearable ultrasonic system. *IEEE Trans Neural Syst Rehabil Eng.* 2014;22(1):69–76.

- 616 42. Akhlaghi N, Baker CA, Lahlou M, Zafar H, Murthy KG, Rangwala HS, et al. Real-time
617 Classification of Hand Motions using Ultrasound Imaging of Forearm Muscles. *IEEE Trans*
618 *Biomed Eng.* 2016;63(8):1687–98.
- 619 43. Baker CA, Akhlaghi N, Rangwala H, Kosecka J, Sikdar S. Real-time, ultrasound-based
620 control of a virtual hand by a trans-radial amputee. In: *Conference Proceedings of the IEEE*
621 *Engineering in Medicine and Biology Society.* 2016. p. 3219–22.
- 622 44. Dhawan AS, Mukherjee B, Patwardhan S, Akhlaghi N, Diao G, Levay G, et al.
623 Proprioceptive Sonomyographic Control: A novel method for intuitive and proportional
624 control of multiple degrees-of-freedom for individuals with upper extremity limb loss. *Sci*
625 *Rep.* 2019;9(1):9499.
- 626 45. Caliński T, Harabasz J. A dendrite method for cluster analysis. *Commun Stat.* 1974;3(1):1–
627 27.
- 628 46. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster
629 analysis. *J Comput Appl Math.* 1987;20:53–65.
- 630 47. Halkidi M, Vazirgiannis M. Clustering validity assessment: finding the optimal partitioning
631 of a data set. In: *Proceedings 2001 IEEE International Conference on Data Mining.* 2001. p.
632 187–94.
- 633 48. Liu Y, Li Z, Xiong H, Gao X, Wu J. Understanding of internal clustering validation
634 measures. In: *2010 IEEE International Conference on Data Mining.* 2010. p. 911–6.
- 635 49. Higgins J. *An Introduction to Modern Nonparametric Statistics.* Pacific Grove, CA:
636 Brooks/Cole; 2004.

- 637 50. National Academies of Sciences, Engineering, and Medicine. The promise of assistive
638 technology to enhance activity and work participation. Jette AM, Spicer CM, Flaubert JL,
639 editors. Washington, DC: National Academies Press; 2017.
- 640 51. Kestner S. Defining the Relationship between Prosthetic Wrist Function and Its Use in
641 Performing Work Tasks and Activities of Daily Living. *J Prosthet Orthot.* 2006;18(3):80–6.
- 642 52. Markovic M, Schweisfurth MA, Engels LF, Farina D, Dosen S. Myocontrol is closed-loop
643 control: incidental feedback is sufficient for scaling the prosthesis force in routine grasping.
644 *J NeuroEngineering Rehabil.* 2018;15(1):81.
- 645 53. Kyranou I, Vijayakumar S, Erden MS. Causes of Performance Degradation in Non-invasive
646 Electromyographic Pattern Recognition in Upper Limb Prostheses. *Front Neurorobotics.*
647 2018;12:58.
- 648 54. Akhlaghi N, Dhawan A, Khan A, Mukherjee B, Diao G, Truong C, et al. Sparsity Analysis
649 of a Sonomyographic Muscle-Computer Interface. *IEEE Trans Biomed Eng.*
650 2019;67(3):688–96.
- 651 55. Hinterreiter A, Ruch P, Stitz H, Ennemoser M, Bernard J, Strobel H, et al. ConfusionFlow:
652 A model-agnostic visualization for temporal analysis of classifier confusion. *IEEE Trans*
653 *Vis Comput Graph.* 2020;
- 654

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
<i>Feedback</i> phase (β_2)	-0.6185	3.0983	-0.1996	0.707
<i>Retention</i> 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)	
Separability	S_Dbw Index	0.483 [†]	-0.259	-0.449	-0.896*	-0.457	-0.563	
		0.099	0.142	0.692	0.262	-0.156 [†]	0.981*	
		-0.123 [†]	0.327	0.509	0.625	0.185	0.849	
	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)	
Separability	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
	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.

Figures

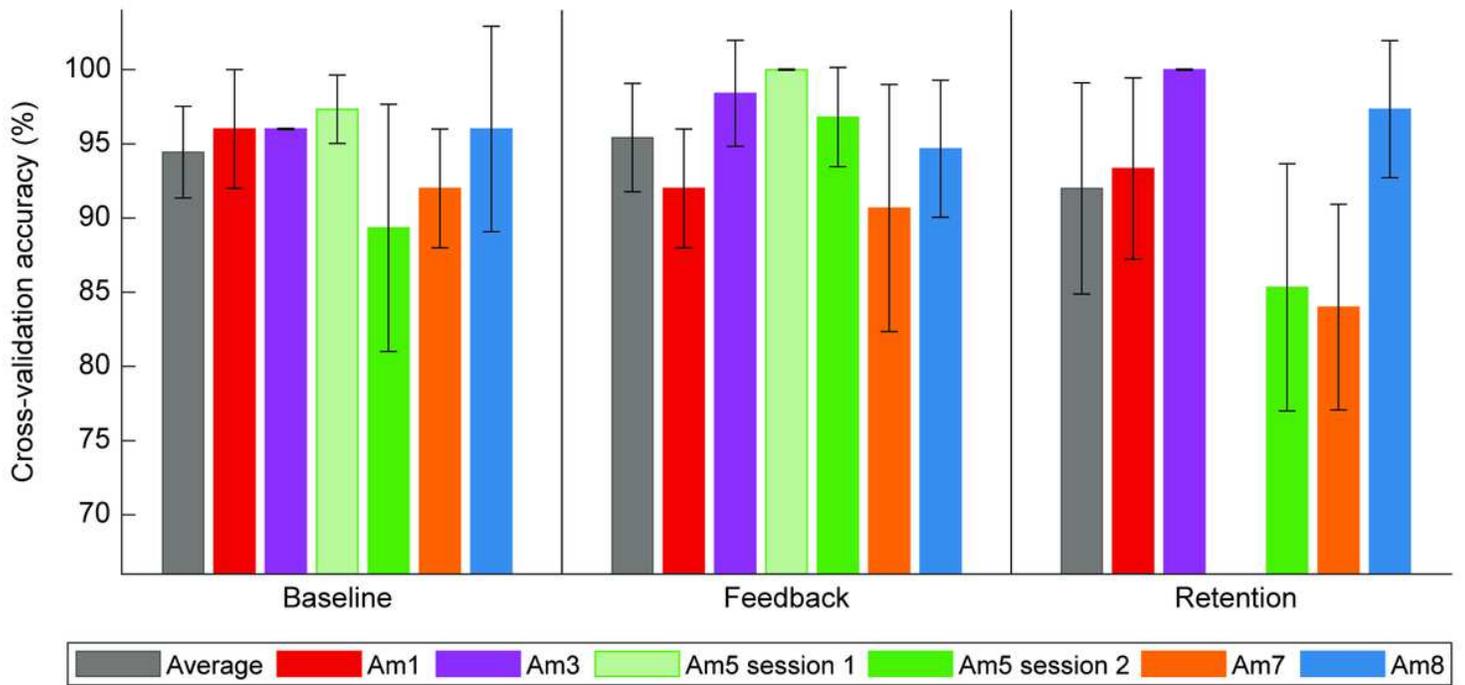


Figure 1

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

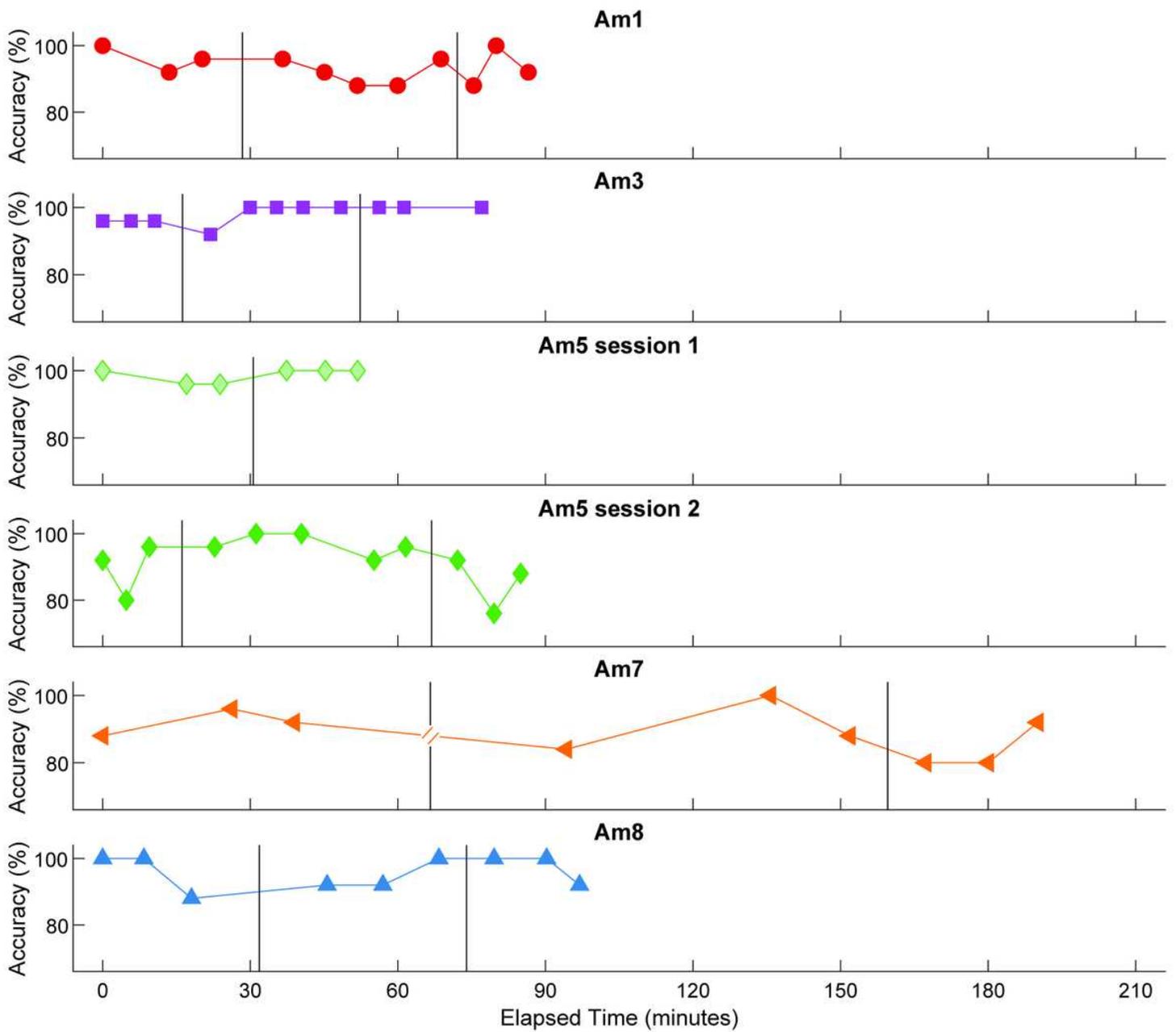


Figure 2

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.

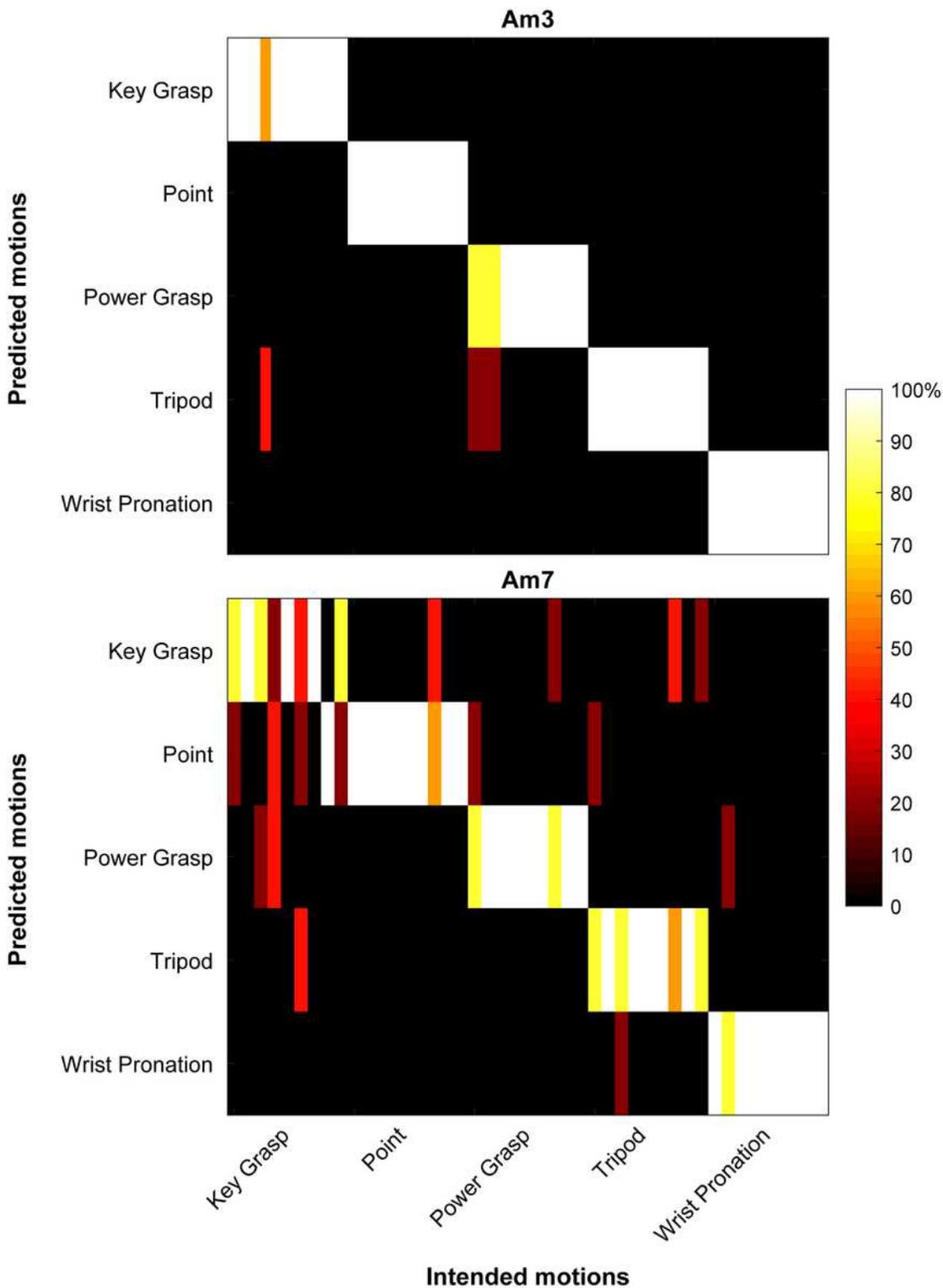


Figure 3

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).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [AdditionalFile3.wmv](#)
- [AdditionalFile2.pdf](#)
- [AdditionalFile1.pdf](#)