

The Influence of Training With Visual Biofeedback on the Predictability of Myoelectric Control Usability

Jena L. Nawfel¹, Kevin B. Englehart¹, *Senior Member, IEEE*, and Erik J. Scheme¹, *Senior Member, IEEE*

Abstract—Studies have shown that closed-loop myoelectric control schemes can lead to changes in user performance and behavior compared to open-loop systems. When users are placed within the control loop, such as during real-time use, they must correct for errors made by the controller and learn what behavior is necessary to produce desired outcomes. Augmented feedback, consequently, has been used to incorporate the user throughout the training process and to facilitate learning. This work explores the effect of visual feedback presented during user training on both the performance and predictability of a myoelectric classification-based control system. Our results suggest that properly designed feedback mechanisms and training tasks can influence the quality of the training data and the ability to predict usability using linear combinations of metrics derived from feature space. Furthermore, our results confirm that the most common in-lab training protocol, screen guided training, may yield training data that are less representative of online use than training protocols that incorporate the user in the loop. These results suggest that training protocols should be designed that better parallel the testing environment to more effectively prepare both the algorithms and users for real-time control.

Index Terms—Electromyography, myoelectric control, user training, online performance, usability, pattern recognition, biofeedback, co-adaptation, linear regression.

I. INTRODUCTION

PATTERN recognition-based myoelectric control has become a popular method used to increase the physical capabilities and independence of functionally-impaired individuals. Assistive devices, such as powered prostheses and orthoses, have leveraged electromyography (EMG) signals from residual functioning muscles to provide users with intuitive control. Still, the abandonment rate of myoelectric prostheses has been reported as 23% [1], with a lack of functionality, insufficient dexterity of control, and unpredictability

Manuscript received August 10, 2021; revised February 24, 2022; accepted March 21, 2022. Date of publication March 25, 2022; date of current version April 6, 2022. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery under Grant 2014-04920 and Grant 217354-15. (Corresponding author: Erik J. Scheme.)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Research Ethics Board of the University of New Brunswick under Approval No. 2020-016.

The authors are with the Institute of Biomedical Engineering, University of New Brunswick, Fredericton, NB E3B 5A3, Canada (e-mail: jnawfel@unb.ca; escheme@unb.ca).

Digital Object Identifier 10.1109/TNSRE.2022.3162421

of the control system being major contributing factors [2]. Several studies have addressed device functionality and control concerns by adjusting the user training protocol to introduce variations into the training data often experienced during real-time use. These confounding factors include variations due to limb position [3]–[6], contraction intensity [6]–[8], and electrode shift [9], [10].

Adaptive myoelectric control also has been investigated to account for variations due to the nonstationarities of the EMG signal. There are two subsystems with adaptive capabilities in a myoelectric control system — the user and the machine. Much focus has been on adapting the machine/classifier by adding new instances of EMG data to the training set using supervised or unsupervised techniques [11]–[13]. The parameters of the classifier are updated based on this new training set to reflect changes in the EMG input signals.

Sensinger *et al.* showed that supervised adaptation tends to outperform unsupervised adaptation, but both paradigms demonstrate reductions in testing error compared to their non-adapting counterparts. Specifically, supervised high-confidence adaptation was recommended as the most practical implementation [11]. This approach reinforces good decisions and accounts for slow drifts in the boundaries of the classifier by appending high-confidence decisions to the pre-existing training set. As more data are added, the classifier is updated so that it may reflect changes in the user's decisions [11]. Another promising solution is supervised low-confidence adaptation. This approach adds low-confidence decisions to the training set resulting in better definition of the classifier boundaries [11]. Other adaptation strategies leverage cycle substitution [13], where for every data point added to the training set, another data point from the same class is removed. Cycle substitution and other methods that dynamically manage the size of the training set reduce the risk of overtraining the classifier [11], [13].

Studies addressing the human adaptation component have found that user practice significantly improves control performance for both proportional control and pattern recognition based control [12], [14]–[17]. Tabor *et al.* revealed that the process of learning myoelectric control may be more gradual than previously assumed. In a ten-day experiment, users reported a perceived performance plateau no earlier than the fifth training session [17]. Furthermore, 87% of participants expected to continue seeing improvements following the 10th

training session had the experiment been extended. He *et al.* discovered that over an eleven day period, users consistently saw improvements in the repeatability of their muscle contractions. The observed performance increase in their study consequently was attributed to the users' ability to learn reproducible EMG patterns [12]. Similarly, it has been shown that users can learn to produce distinguishable muscle patterns for different motion classes. Bunderson *et al.* found that the separability index is significantly higher in experienced myoelectric users than in novice users, suggesting that practice can improve this metric and lead to better performance [16]. Users have even demonstrated the ability to learn non-intuitive myoelectric control schemes with adequate training [14]. Over time, user performance with a non-intuitive interface converged to be consistent with that of an intuitive one [14]. This evidence highlights the human capacity to learn complex and abstract tasks.

The majority of studies assessing adaptation have focused on either machine adaptation or user adaptation. Only recently have researchers begun to implement closed-loop myoelectric training schemes where both learners, human and machine, work together to converge to a common solution [18]–[20]. The literature has termed the cooperation between user and machine, and how it evolves over time, as co-adaptation. In this scenario, both user and machine are learning while focusing on a common goal — to maximize control performance [19].

For successful co-adaptation, a bidirectional communication channel must be established between the two learners [21]. In one direction, myoelectric signals that describe movement intent are sent from the human to the machine. This feed-forward channel encourages machine learning and is well established in the literature [6]. In the other direction, the machine's interpretation of the incoming data is provided to the human from the machine, promoting the user's understanding of the control system [22]. The most commonly cited strategy for training a myoelectric control system (i.e., screen guided training) only makes use of the feed-forward communication channel [11], [23]–[25]. However, relaying information through the feedback channel allows users to regulate their muscle contractions with knowledge of how the classifier responds, which could potentially lead to improved control during real-time use [19]. This type of augmented feedback has been shown to improve both understanding and control of human-machine interfaces [22], [26]–[28].

Visual augmented feedback has been investigated extensively to help users modulate their level of muscle activity, improve muscle coordination, and reduce classification error [29]–[34]. Incorporating visual feedback during user training also has become a popular method to facilitate user learning in both sequential and simultaneous control schemes [12], [15], [16], [23], [24], [35]–[39]. Studies focusing on providing classification-based feedback for sequential control have found that it can improve both short-term and long-term user performance [15], [23], [24], [35]. Classification-based feedback presents users with information about the patterns and the cumulative effect of their movements in the context of a classification scheme.

This allows users to be aware of the misclassifications made by the system and respond to them accordingly.

Works investigating simultaneous control schemes have supported the training process with visual feedback to help users better match their target trajectories [37], [38]. Ison *et al.* and Hahne *et al.* implemented feedback that displayed both the target wrist position and the actual position obtained by a motion tracking system on a computer screen to generate more robust input mappings [37], [38].

Other studies have focused on presenting users with feedback that monitors the spatial distribution of the EMG signal using polar plots [24], [40]. The feedback is used to ensure 1) that each motion is distinct and 2) that the patterns made for a particular gesture are repeatable. Roche *et al.* suggests that reinforcement of hand movements using these polar plots, in combination with imitation and repetition, improves the control of a multi-functional prosthetic hand in just one session [40].

Clustering-based feedback is another strategy that has been proposed in the literature [41]. This approach provides users with real-time visual feedback about their online EMG signals as well as with the centroids of their previous training class clusters. The feedback allows users to focus on the repeatability of their EMG signals and adjust their muscle contractions to better reflect their previous training instances. A study comparing classification-based feedback and clustering-based feedback found that the continuous nature of the clustering approach was more beneficial and led to higher classification accuracies than discrete decisions provided by the classifier [41]. These results suggest that providing feedback associated with the underlying muscle activation patterns may encourage a more comprehensive understanding of what is necessary for successful pattern recognition control [22].

Studies also have demonstrated that training with different levels of visual augmented feedback influences the quality of the training data [16], [24], [35]. Training sessions with no feedback motivate the user to adopt an internal focus during movement execution [24], [42]. When visual feedback is provided, the users' focus is shifted to the external source of information. This not only changes the users' attentional focus, but also their behavior, and consequently their EMG signals [24], [42]. Kristoffersen *et al.* found in their study that the amplitude of the EMG signal and the variability of muscle patterns tend to be lower when training with visual feedback [24]. These results support earlier findings that directing the users' attention to the effect of their movements, rather than on the actual movements themselves, leads to lower levels of EMG activity and a greater degree of automaticity [42].

Implementing appropriate visual feedback in a co-adaptive training strategy may provide the necessary variations in the training data to address concerns regarding the predictability of the control system. For years, researchers have used offline training metrics, such as offline classification accuracy, as a proxy for real-time performance. However, more recent evidence suggests that many offline metrics are poorly correlated with online performance [15], [43]–[46]. Our previous work aimed to bridge this information divide by investigating the

relationship between training data characteristics and usability for a classification-based myoelectric control scheme [47]. We identified offline metrics that provided better predictions of usability than the current standards in the literature (i.e., classification accuracy, separability index, repeatability index), and found that a combination of offline metrics may provide a better indication of real-time performance compared to individual metrics. The following study expands on that previous work by influencing the training protocol using a feedback channel that provides users with visual information regarding their muscle activation patterns. In addition to testing screen guided training, we implemented two categories of visual EMG biofeedback (repeatability-based feedback and separability-based feedback) and two versions of visual functional classifier-based feedback (constant velocity and proportional control). If differences in user behavior due to the observed feedback manifest themselves in the training data, then the resulting feature space metrics may provide better predictions of real-time performance.

II. METHODS

A. Participants

Twelve normally-limbed individuals (9 male/3 female, age range: 22-63 yrs., mean and standard deviation of age: 33 ± 15.2 yrs., median age: 25.5 yrs.) participated in this experiment. Ten subjects reported right hand dominance and two reported left hand dominance. The procedures were approved by the research ethics board at the University of New Brunswick (REB #2020-016), and subjects provided written informed consent prior to their participation in the study.

B. Experimental Setup

A non-invasive, wireless EMG collection system (Trigno™ Wireless system, Delsys Inc., USA) with a sampling frequency of 2000 Hz was used to record all EMG signals. The signals were filtered using 2nd-order Butterworth bandstop filters at 60, 180, 250, and 300 Hz to remove power-line and digital interference. A 3rd-order Butterworth high-pass filter with cutoff frequency of 20 Hz was also implemented to address motion artifact. Before positioning the electrodes on the subjects, the proximal third of the dominant forearm was cleansed with an alcohol swab. Six EMG electrodes were then uniformly spaced around the forearm in the area with the largest muscle bulk. During experimentation, participants were seated with their dominant arm resting comfortably at a 90-degree angle.

C. Experimental Protocol

The experiment consisted of six separate sessions, each conducted on a different day. The first session allowed subjects to practice using each feedback modality in a 45-minute familiarization experience. Each of the remaining five experimental sessions lasted about 20 minutes and consisted of a specific training protocol (as described below in Section II-C.1), followed by a testing phase to evaluate classifier usability. A within-subject design was chosen to reduce the presence of

inter-subject confounding factors, and the order of the feedback training methods was randomized to mitigate learning effects. EMG signals for five motion classes were collected: no movement, wrist flexion, wrist extension, power grip, and hand open. Each cycle through these five movements constituted a trial, and the order that these movements were presented to the user remained fixed.

1) **Feedback:** A custom graphical user interface (GUI) developed in MATLAB® was used to introduce four feedback modalities during training (Figure 1): screen guided training, repeatability-based feedback, separability-based feedback, and functional-based feedback.

The screen guided training protocol presents no information back to the user about the behavior of the control system. In contrast, the functional-based feedback mimics the testing environment and provides users with direct information about how their muscle contractions affect the classifier's decisions in real-time. The repeatability- and separability-based feedback methods incorporate the user in the control loop, but the information conveyed to users is not directly indicative of the classifier's decisions. Instead, these types of feedback provide information about how the users' muscle contractions affect feature space characteristics.

When users are presented with relevant feedback about their muscle activity, they have the ability to modulate their movements and contraction intensity in real-time. Therefore, it is important that the feedback is representative of high-quality movements so that users may continue to perform their contractions properly. In this study, the data used for visualization purposes was based on an adaptive classifier's ability to correctly predict new instances of a movement.

The two continuous feedback modalities (i.e., repeatability- and separability-based feedback) displayed only windows of data that were both correctly classified and collected during the last two seconds of a repetition. Because these modalities used position control (the cursor location on the screen was proportional to the contraction intensity), displaying the second half of the previous repetitions helped to ensure that the data used for visualization were targeting where users should end up with their contractions and not the path they took to get there. A forgetting factor of four trials was implemented so that data from at most four previous trials were included in the visualization.

Because the functional-based feedback displayed task-focused classification decisions, the location of the cursor was updated based on the decisions of the existing classifier. The size and location of the target for functional-based feedback were not dependent on the collected data; therefore, these feedback modalities employed velocity control (the velocity of the cursor was proportional to the contraction intensity), rather than positional control. For this reason, there were no constraints placed on the portion of data used for visualization. Detailed descriptions of each feedback condition are provided below.

- **Screen Guided Training (SGT):** Screen guided training [48], shown in Figure 1a, prompted the user to perform a given movement using a visual cue. An image of a hand gesture was shown on the screen with a

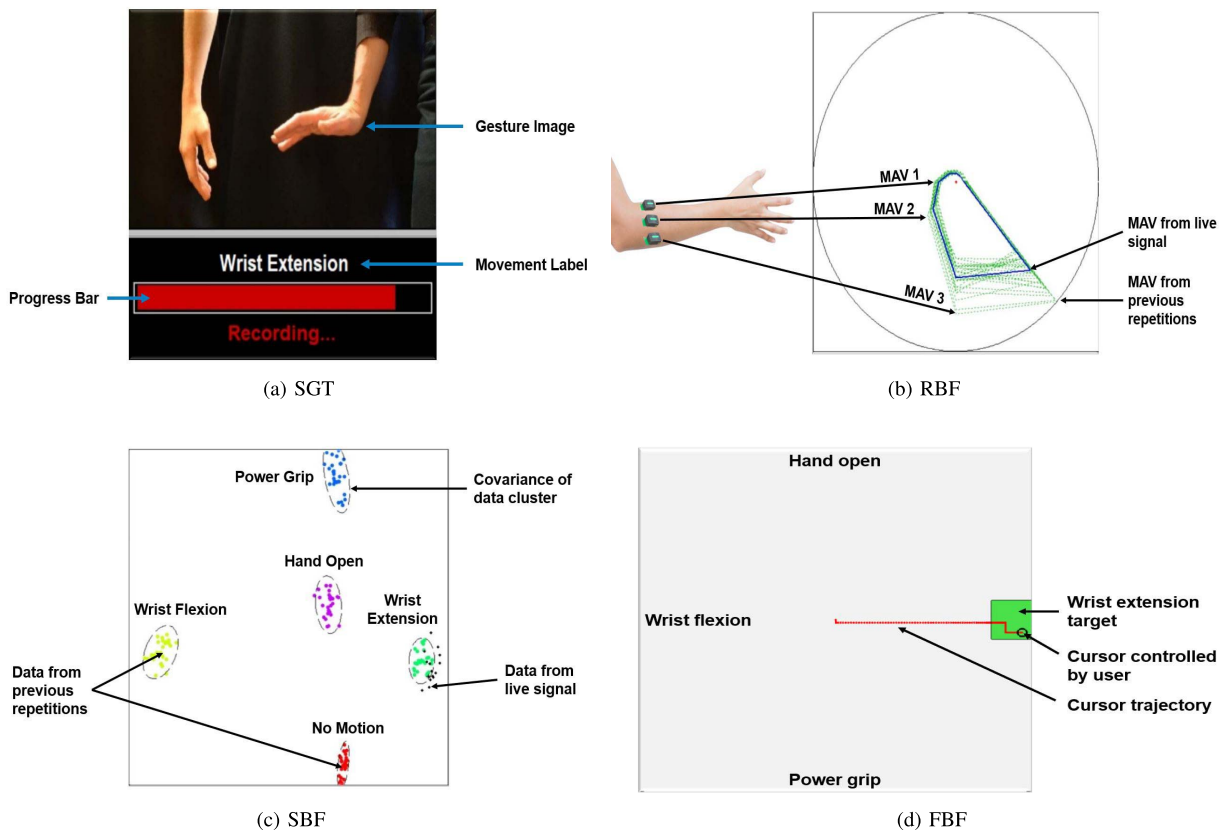


Fig. 1. Illustrations of the feedback modalities presented to users during training.

progress bar informing the user how long to maintain their contraction. Subjects began each movement at rest, transitioned into the desired movement, and then held the contraction for the duration of the repetition. The system recorded four seconds of EMG data for each prompt followed by a two-second delay period during which no data were recorded. The delay allowed users to return to a resting position before the next prompted movement.

- Repeatability-Based Feedback (RBF) + SGT:** Repeatability-based feedback displayed the live EMG signal as a polygon where each vertex represented the mean absolute value (MAV) of the signal at the corresponding electrode channel. Static dotted-lined polygons representing the channel-by-channel MAV of previous repetitions for a given motion class were plotted behind the live signal. The goal was to position the live polygon signal over the areas with the highest density of static polygons. Operant conditioning turned the color of the polygons green if the MAV of the live EMG signal at each electrode channel was within the bounds of the static polygons, as shown in Figure 1b. The inner and outer bounds were computed as the minimum and maximum MAV values observed at each channel over the previous repetitions.
- Separability-Based Feedback (SBF) + SGT:** The separability-based feedback in Figure 1c provided users with information about muscle pattern repeatability and separability. The four time domain features described by

Hudgins *et al.* [49] were extracted from the EMG signal at each electrode channel for a total of 24 features. An uncorrelated linear discriminant analysis (uLDA) projection was performed on the features to reduce the dimensionality to two dimensions. The projection was then displayed for users on the computer screen. The goal of this feedback was to position the projected live signal over the cluster of data points corresponding to the prompted motion class.

- Functional-Based Feedback (FBF):** Functional-based feedback provided users with task-focused classification decisions associated with their contractions. A two degree of freedom (DOF) Fitts' Law environment, like the one shown in Figure 1d, was displayed on a computer screen. The four active hand motions were mapped to control the movement of the cursor. In the vertical DOF, hand open moved the cursor up and power grip moved the cursor down. In the horizontal DOF, the direction of wrist extension and wrist flexion was mapped to match that of the subject's reported hand dominance. Targets were positioned in one DOF axis only, meaning a single active hand movement was sufficient to acquire the target. If misclassifications were observed, additional compensatory movements may have been necessary to successfully complete the trial. Users were given a total of ten seconds to position the cursor in the correct location on the screen. Once the cursor reached the target, it was required to stay within its bounds for a dwell time of

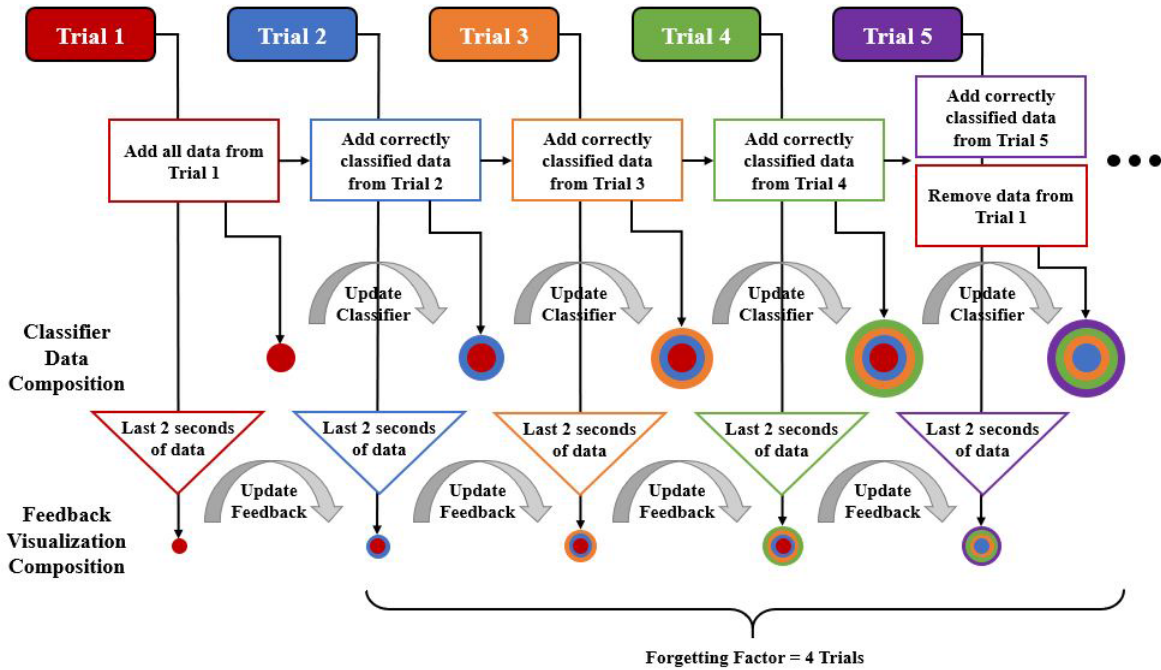


Fig. 2. The procedure for updating the visualizations during the continuous feedback modalities.

one second. Two types of functional-based feedback were tested. One with constant velocity control (FBF_{cv}) and the other with proportional control (FBF_{pc}). During constant velocity control (i.e. “on/off” control), the travel speed of the cursor remained constant throughout user training when an active class was determined. During proportional control, the travel speed of the cursor was proportional to the user’s contraction intensity, allowing the user to have more control over the target acquisition task.

2) Classification Scheme: This experiment employed an adaptive linear discriminant analysis (LDA) classifier. An initial classifier was trained following the completion of the first trial. After every additional trial, the LDA classifier was retrained, taking into account data from the previous trial. A forgetting factor of four trials was implemented so that the classifier was trained using only the four most recent trials, as shown in Figure 2. For more details on the adaptation scheme, please refer to [47].

The classifiers were trained using the four time domain features described by Hudgins *et al.* (mean absolute value, slope sign changes, zero crossings, and waveform length) [49]. The training data was segmented into 160 ms windows with a 64 ms overlap.

3) Familiarization Session: The familiarization session encouraged users to interact with the interface and to gain knowledge and understanding about the various feedback modalities. The five forms of feedback (i.e., SGT, RBF, SBF, FBF_{cv} , and FBF_{pc}) were presented and explained to users in a randomized order. Six practice trials were collected for each feedback type.

During the first trial, SGT prompted the user to perform the set of movements. If the selected feedback was SGT, then the remaining trials were identical to the first trial. Otherwise,

the subsequent trial displayed visual information to the user based on the data acquired during the first SGT trial. As more EMG data were collected, the visual feedback was updated to incorporate the new muscle patterns from previous trials. The algorithm had a forgetting factor of four trials; therefore, the visual feedback presented during the sixth trial was based on the data from the four most recent trials (i.e., trials 2–5). The intent was that this initial exposure to the feedback prior to experimentation would help minimize the associated learning effect.

4) Experimental Sessions: The experimental sessions tested each feedback modality separately. The feedback testing order was randomized across subjects to reduce the impact of potential learning effects. Each session was comprised of a training and testing portion.

Part 1 (Training): The training portion of the study consisted of eight training repetitions of each motion class. The feedback was presented in the same way as during the familiarization session — where the first trial was acquired with SGT, and the remaining seven trials were collected while displaying the selected feedback modality. Again, the feedback shown during any given trial was dependent on the data acquired during the previous trials with a forgetting factor of four trials. The adaptive classifier also had a forgetting factor of four trials; by the end of training, the classifier contained correctly classified decisions from trials 5–8.

Part 2 (Testing): Following the feedback training session, subjects tested the usability of the classifier in a Fitts’ Law target acquisition task. The movement of the cursor in the virtual environment was controlled by the users’ muscle activity. The control of the cursor was the same way as in the FBF training protocol. The vertical DOF was controlled by hand open and power grip, and the horizontal DOF was controlled by wrist flexion and wrist extension. Proportional control was also

provided so that the subjects' contraction intensity regulated the speed of the cursor. Unlike the FBF training, the Fitts' Law testing phase involved targets at different locations in virtual space and targets that required activation of one or both DOFs. The user successfully completed a test trial after acquiring the target and maintaining the cursor within its boundaries for a dwell time of one second. The timeout for each trial was ten seconds, after which the test automatically advanced to the next target. The testing phase included a total of 64 targets, half of which were single DOF targets (e.g. right or down) and the other half dual DOF targets (e.g., left AND up). New windows were extracted every 16 ms using the most recent 160 ms of data. From these windows, features were extracted and passed to the LDA classifier where a decision was made and mapped to be visually observed by users in the Fitts' Law environment.

D. Feature Space Metrics

In our previous work, we extracted a total of 32 offline training metrics from the EMG feature space [47]. The current study leverages all of these same measures with the addition of a training dependent measure. This not only facilitated comparisons in performance for the different training schemes, but it also allowed the predictive models to be developed analogously to those in [47]. The training dependent measure was used to characterize user behavior during functional-based feedback. The amount of training time (TT) for the functional-based feedback is dependent on how well the user can reach the training targets. Users that had difficulty acquiring the targets will naturally be required to train longer and have more training data than users that acquire the targets easily. The other three feedback modalities had a relatively constant training time because the user was prompted with movements for a designated amount of time. This added metric, therefore, was intended for the functional-based feedback cases and is defined as,

$$TT = (n) (win_{size}) - (win_{inc})(n - 1) \quad (1)$$

where n is the total number of data windows in the training session, win_{size} is the size of the window in ms, and win_{inc} is the amount of time between subsequent windows in ms.

Table I provides brief descriptions and abbreviations for the 33 feature space metrics. For mathematical formulas, please refer to [47] or the inline citations in Table I.

As in [47], the analysis leveraged the full training repetitions rather than only the portions applied to the adaptive classifier to fully evaluate the user's behavior throughout the training process.

E. Online Performance

Metrics Fitts' Law style testing has become an increasingly popular method to evaluate myoelectric control [7], [55], [56]. The following online performance metrics were extracted from the Fitts' Law testing environment and are cited as some of the more commonly used virtual outcomes in the literature [7], [57], [58]: throughput (TP), effective throughput (eTP), path efficiency (PE), overshoot (OS), average speed (AS), and stopping distance (SD).

F. Correlation Analysis

A correlation analysis was performed to measure the strength of association between each feature space and online performance metric across feedback conditions. Normality was assessed using the Shapiro-Wilk normality test. The Pearson correlation coefficient was calculated between the corresponding metrics when the assumption of normality held. When the assumption of normality was violated, Kendall's coefficient of rank correlation was implemented [59].

G. Linear Regression

The relationship between training data characteristics and online myoelectric control was evaluated using the feature space metrics as predictors and the online performance metrics as response variables. An ordinary least squares regression approach was implemented to quantify the relationship on a given response variable using sets of one, two, and three predictor variables. Both predictors and responses were normalized prior to model development.

1) *Predictor Selection*: Our predictor selection approach was based on the consensus nested cross-validation technique proposed by Parvande *et al.* for feature selection [60]. This approach is designed to find consistent and stable features, while encouraging the formation of a generalizable model [60]. For a detailed description of our implementation procedures, please refer to [47].

2) *Performance Evaluation*: The predictive performance of each selected model based on the predictor selection procedures was determined using a leave-one-subject-out (LOSO) cross validation across all twelve subjects. The predictor set for the one-, two-, and three-predictor models demonstrating the lowest predictive mean squared error was selected as having the "best" performance. The predictive accuracy of these final models was evaluated using the variance explained (VEcv) metric described by Li [61], [62]. The VEcv metric is sometimes referred to as the predicted R^2 and is defined as,

$$VEcv = \left(1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \right) \times 100 \quad (2)$$

where n is the number of samples, y is the true value, \hat{y} is the predicted value, and \bar{y} is the mean of the validation data. Negative VEcv values indicate that the model predictions are less accurate than predicting using the mean of the validation data [61]. Positive VEcv values demonstrate that the model predictions are more accurate than using the validation mean [61]. According to Li, the performance of predictive models based on VEcv measures can be divided into the following five categories: *very poor*: $VEcv \leq 10\%$, *poor*: $10\% < VEcv \leq 30\%$, *average*: $30 < VEcv \leq 50\%$, *good*: $50 < VEcv \leq 80\%$, *excellent*: $VEcv > 80\%$.

H. Statistical Analysis

The main effects of session and feedback on the outcome variables were examined using either a repeated measures analysis of variance (RM-ANOVA) or a non-parametric Friedman's test. Due to sample size limitations, the number of participants was insufficient to support a model

TABLE I
FEATURE SPACE METRIC DESCRIPTIONS. ADAPTED FROM APPENDIX A IN [47]

Metric	Abbrev.	Description
Repeatability Index	RI	A measure of the reproducibility of EMG patterns between repetitions [16]
Mean Within-Repetition Repeatability Index	mwRI	A measure of the reproducibility of EMG patterns within a single repetition [47]
Std. Dev. Within-Repetition Repeatability Index	swRI	A measure of the variation of the within-repetition repeatability across repetitions [47]
Std. Dev. Within-Trial Separability Index	swSI	A measure of the variability of the distinguishability of EMG patterns across trials [47]
Mean Semi-Principal Axes	MSA	A measure of intraclass variability that quantifies the size of a training ellipsoid [16]
Centroid Drift	CD	A measure that quantifies the variation in centroid location of a training ellipsoid across subsequent repetitions [47]
Mean Absolute Value	MAV	A measure that specifies the average amplitude of the EMG signal [49]
Separability Index	SI	A measure of interclass distance [16]
Modified Separability Index	mSI	A measure similar to the separability index, except that it accounts for the covariance matrix of both distributions being compared [46]
Mean Within-Trial Separability Index	mwSI	A measure of the distinguishability of EMG patterns within a trial [47]
Bhattacharyya Distance	BD	A measure of the statistical similarity between two distributions [50]
Kullback-Leibler Divergence	KLD	A measure of how well a distribution can be approximated by a reference distribution [46]
Hellinger Distance Squared	HD	A measure that quantifies the similarity between two probability distributions [46]
Fisher's Discriminant Ratio	FDR	A measure of separability based on maximizing the inter-class variance and minimizing the intra-class variance
Volume of Overlap Region	VOR	A measure of the degree of overlap between the tails of two class conditional distributions [51]
Feature Efficiency	FE	A measure of the fraction of points separable by a particular feature [51]
Trace of Scatter Matrices	TSM	A measure of class discriminability based on the within-class and between-class scatter matrices [52]
Desirability Score	DS	A function of the separability index, the mean semi-principal axes, and the repeatability index [53]
Class Discriminability Measure	CDM	A measure describing the relationship between clusters in feature space [52]
Purity	PU	A measure assessing the homogeneity of the training data at different resolutions [54]
Rescaled Purity	rPU	A variation of the purity metric weighted and rescaled as described in [54]
Neighborhood Separability	NS	A measure quantifying the relationship between nearest neighbors [54]
Rescaled Neighborhood Separability	rNS	A variation of the neighborhood separability metric weighted and rescaled as described in [54]
Collective Entropy	CE	A measure representing the accumulated uncertainty in the data across different resolutions [54]
Rescaled Collective Entropy	rCE	A variation of the collective entropy metric weighted and rescaled as described in [54]
Compactness	C	A measure estimating the spread of the training data
Rescaled Compactness	rC	A variation of the compactness metric weighted and rescaled as described in [54]
Classification Accuracy	CA	A measure describing the fraction of predictions the classifier labelled correctly [8]
Active Classification Accuracy	ACA	A measure describing the fraction of predictions the classifier labelled correctly excluding misclassifications due to no motion [8]
Usable Data	UD	A measure describing the percentage of correctly classified decisions over the entire user training period using an adaptive classifier [47]
Inter-Class Fraction	ICF	A measure describing the ratio of the number of inter-class nearest neighbors to the total number of samples in the data set [51]
Intra-Inter Fraction	ICF	A measure describing the ratio of the average euclidean distance of intra-class and inter-class nearest neighbors [51]
Training Time	TT	A measure quantifying the total amount of training time

assessing the interaction between session and feedback. For each feature space metric and each online performance metric, the Shapiro-Wilk normality test was conducted across feedback conditions and across sessions to test whether

the data in each group followed a normal distribution. A RM-ANOVA was used if the assumption of normality failed to be rejected and a Friedman's test was used if any one level of the variable dissatisfied the normality

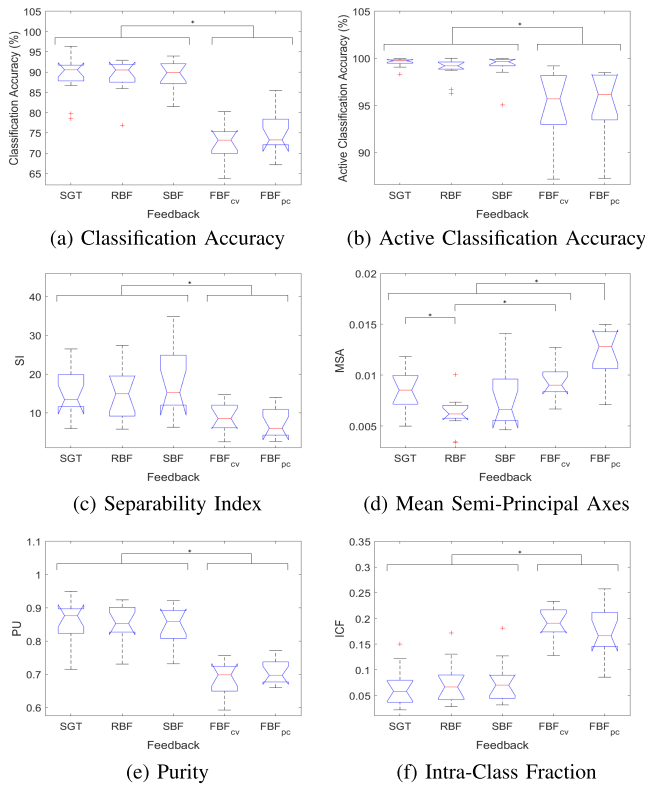


Fig. 3. Boxplots describing the distribution of select feature space metrics for each feedback modality: screen guided training (SGT), repeatability-based feedback (RBF), separability-based feedback (SBF), functional-based feedback with constant velocity control (FBF_{cv}), and functional-based feedback with proportional control (FBF_{pc}). An asterisk indicates statistical significance at the 95% confidence level. The two functional-based feedback approaches generally led to significantly “worse” feature space metrics.

assumption. The Greenhouse–Geisser correction was applied to the RM-ANOVA when Mauchly’s test of sphericity indicated a violation. Post hoc analysis involved multiple comparison tests using a Bonferroni correction with a significance level of $\alpha = 0.05$. Each statistical family consisted of five comparisons corresponding to either the number of feedback options or the number of sessions.

III. RESULTS

A. Training Performance

The distributions of selected feature space metrics across subjects and feedback modalities are shown in Figure 3. Many of the feature space metrics showed similar trends, where the FBF protocols yielded significantly different results than the other three feedback conditions.

Non-parametric Friedman’s tests rendered significant results indicating an overall difference among feedback conditions for classification accuracy (CA) ($\chi^2(4) = 32.7$, $p < 0.001$) and active classification accuracy (ACA) ($\chi^2(4) = 33.8$, $p < 0.001$). Post hoc comparisons using the Bonferroni correction revealed that FBF_{cv} produced significantly lower CA and ACA than SGT ($p_{CA} < 0.001$, $p_{ACA} < 0.001$), RBF ($p_{CA} < 0.001$, $p_{ACA} = 0.019$), and SBF ($p_{CA} < 0.001$, $p_{ACA} < 0.001$). FBF_{pc} also produced significantly lower CA and ACA than SGT

($p_{CA} = 0.013$, $p_{ACA} = 0.001$), RBF ($p_{CA} = 0.013$, $p_{ACA} = 0.044$), and SBF ($p_{CA} = 0.008$, $p_{ACA} < 0.001$).

The SBF training scheme resulted in the highest average separability index (SI) across subjects and demonstrated significant differences compared to the two FBF training strategies ($p_{FBF_{cv}} = 0.006$, $p_{FBF_{pc}} = 0.008$). SGT and RBF also yielded significantly higher SI than the FBF_{cv} ($p_{SGT} = 0.013$, $p_{RBF} = 0.006$) and FBF_{pc} ($p_{SGT} = 0.002$, $p_{RBF} = 0.021$) protocols.

Additionally, the mean semi-principal axes (MSA) demonstrated significant differences across the feedback modalities. The lowest MSA was observed during the RBF training scheme, and significant differences were observed between RBF and SGT, FBF_{cv}, and FBF_{pc} ($p = 0.047$, $p < 0.001$, $p < 0.001$, respectively). SGT, SBF, and FBF_{cv} also exhibited significantly lower MSA than FBF_{pc} ($p = 0.029$, $p = 0.026$, $p = 0.003$, respectively).

The purity metric (PU) and intra-class fraction (ICF) also rendered significant differences across the FBF sessions and the SGT, RBF, and SBF sessions, as shown in Figures 3e and 3f, respectively. The repeatability index (RI) is another commonly cited feature space metric. Although not plotted here, we observed no significant differences for the RI across the different feedback modalities.

The effect of session on the feature space metrics was investigated to determine whether learning effects confounded the results. All statistical tests rendered non-significant results indicating that there were no observed session effects across the feature space metrics.

B. Online Performance

The distributions of the six online performance metrics across feedback conditions and sessions are shown in Figures 4 and 5, respectively. Interestingly, while the functional-based feedback approaches generally led to a significant degradation in the feature space metrics, there was generally no observable decrease in the online performance metrics. In fact, in some cases, the functional-based feedback actually led to marginal improvements in online performance for path efficiency, overshoot, and stopping distance. These results suggest that the relationship between feature space metrics and online performance metrics may not be a direct one. Statistical differences in the online performance metrics across feedback conditions and sessions are described in the text below.

1) **Throughput:** A RM-ANOVA was conducted to determine whether a difference existed between the mean throughput values obtained after training with the different feedback conditions ($F(4,44) = 1.42$, $p = 0.24$). The results identified no significant differences in throughput between SGT, RBF, SBF, FBF_{cv}, and FBF_{pc}.

The main session effect was evaluated using a non-parametric Friedman’s test. The test indicated that throughput was significantly different across sessions ($\chi^2(4) = 18$, $p = 0.001$), suggesting the potential presence of a learning effect. Post hoc comparisons using the Bonferroni correction revealed that session 5 produced significantly higher throughput values

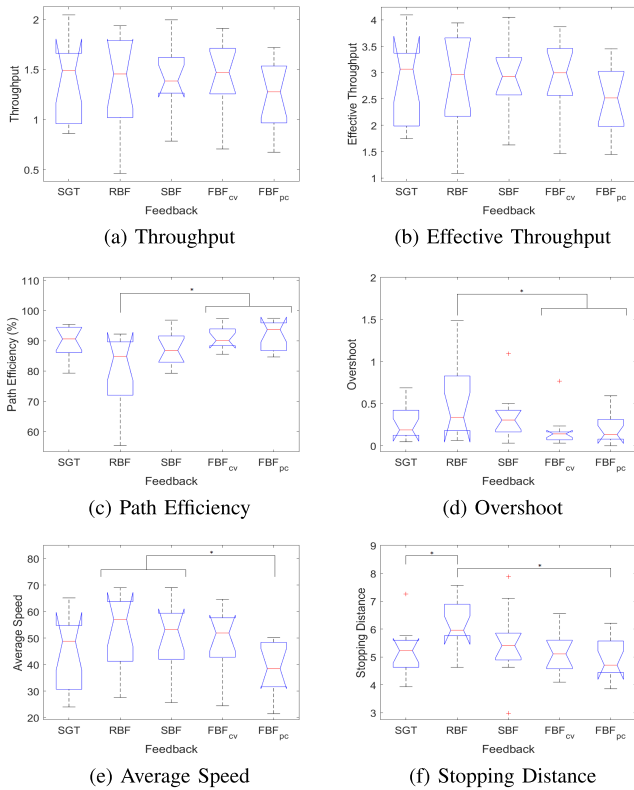


Fig. 4. Boxplots describing the distribution of the online performance metrics for each feedback modality: screen guided training (SGT), repeatability-based feedback (RBF), separability-based feedback (SBF), functional-based feedback with constant velocity control (FBF_{cv}), and functional-based feedback with proportional control (FBF_{pc}). An asterisk indicates statistical significance at the 95% confidence level. The investigated feedback conditions (RBF, SBF, FBF_{cv}, and FBF_{pc}) generally did not show statistical differences in the online performance metrics compared to the control condition (SGT).

compared to sessions 1 ($p = 0.003$) and 2 ($p = 0.045$), and session 4 produced significantly higher throughput than session 1 ($p = 0.045$). No significant differences were found for session 3.

2) Effective Throughput: A RM-ANOVA was conducted to determine whether a difference existed between the mean effective throughput values across the different feedback conditions. The results ($F(4,44) = 1.87$, $p = 0.13$) identified no significant differences in effective throughput between SGT, RBF, SBF, FBF_{cv}, and FBF_{pc}.

The main session effect was evaluated using a RM-ANOVA ($F(4,44) = 8.09$, $p < 0.001$). Post hoc comparisons using the Bonferroni correction revealed that session 1 produced significantly lower throughput values compared to session 3 ($p = 0.011$), session 4 ($p = 0.047$), and session 5 ($p = 0.011$). No significant differences were observed for session 2.

3) Path Efficiency: A non-parametric Friedman's test was conducted to determine whether differences existed between the path efficiency values across the feedback conditions ($\chi^2(4) = 16.8$, $p = 0.002$). The significant result indicated the presence of an overall difference among feedback conditions. Post hoc comparisons using the Bonferroni correction revealed that FBF_{cv} ($p = 0.020$) and FBF_{pc} ($p = 0.003$) produced

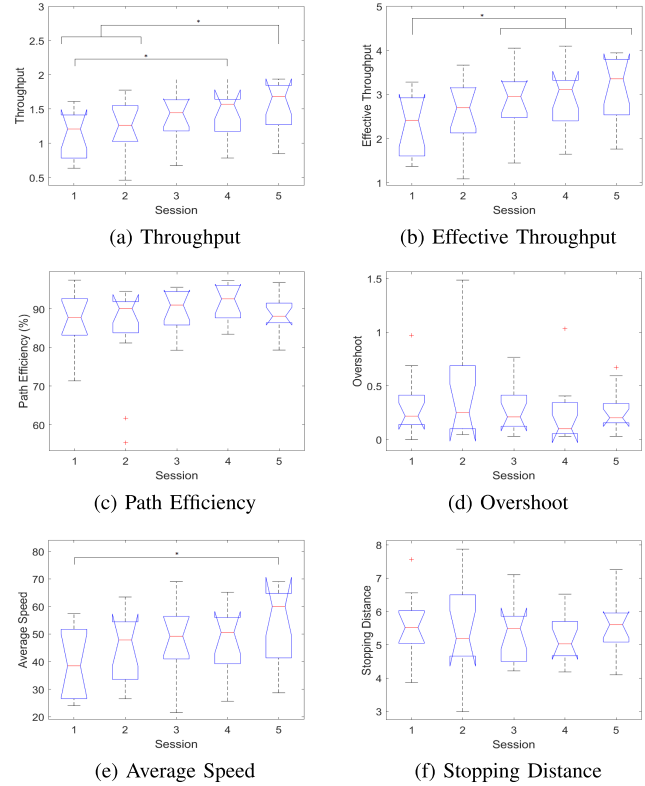


Fig. 5. Boxplots describing the distribution of the online performance metrics for each session. An asterisk indicates statistical significance at the 95% confidence level. A significant session effect was observed for throughput, effective throughput, and average speed, indicating the presence of a potential learning effect.

significantly higher path efficiency than RBF. No significant differences were found for SGT or SBF.

The main session effect was evaluated using a non-parametric Friedman's test ($\chi^2(4) = 3.27$, $p = 0.51$). The results identified no significant differences in path efficiency values between session 1, session 2, session 3, session 4, and session 5.

4) Overshoot: A non-parametric Friedman's test was conducted to determine whether differences existed between the overshoot values across the different feedback conditions ($\chi^2(4) = 14.2$, $p = 0.006$). Post hoc comparisons using the Bonferroni correction revealed that FBF_{cv} ($p = 0.012$) and FBF_{pc} ($p = 0.035$) produced significantly lower overshoot than RBF. No significant differences were found for SGT or SBF.

The main session effect was evaluated using a non-parametric Friedman's test ($\chi^2(4) = 4.62$, $p = 0.33$). The results identified no significant differences in overshoot values between session 1, session 2, session 3, session 4, and session 5.

5) Average Speed: A RM-ANOVA was conducted to determine whether a difference existed between the mean average speed values across the different feedback conditions ($F(4,44) = 5.84$, $p < 0.001$). Post hoc comparisons using the Bonferroni correction revealed that FBF_{pc} produced significantly lower average speed than RBF ($p = 0.004$) and SBF

($p = 0.020$). No significant differences were observed for SGT or FBF_{cv} .

The main session effect was evaluated using a non-parametric Friedman's test. The test indicated that average speed was significantly different across sessions ($\chi^2(4) = 14.33$, $p = 0.006$). Post hoc comparisons using the Bonferroni correction revealed that session 5 produced significantly higher average speed compared to session 1 ($p = 0.002$). No significant differences were found for session 2, session 3, or session 4.

6) Stopping Distance: A RM-ANOVA was conducted to determine whether a difference existed in the mean stopping distance across the different feedback conditions. The test indicated that stopping distance was significantly different across the types of feedback ($F(2,2,24.3) = 4.03$, $p = 0.03$). Post hoc comparisons using the Bonferroni correction revealed that FBF_{pc} ($p = 0.008$) and SGT ($p = 0.02$) produced significantly lower stopping distance compared to RBF. No significant differences were found for SBF or FBF_{cv} .

The main session effect was evaluated using a RM-ANOVA ($F(4,44) = 0.34$, $p = 0.85$). The results identified no significant differences in stopping distance between session 1, session 2, session 3, session 4, or session 5.

C. Correlation Analysis

Summaries of the correlations for the five EMG feedback modalities are described below.

1) Screen Guided Training: Feature efficiency (FE) demonstrated the strongest significant correlation with throughput ($r = 0.515$, $p = 0.021$) and effective throughput ($r = 0.485$, $p = 0.031$). Path efficiency was most correlated with the Bhattacharyya distance (BD), ($r = -0.595$, $p = 0.041$) while average speed was most correlated with the rescaled purity metric (rPU), ($r = 0.623$, $p = 0.030$). Stopping distance was most correlated with Fisher's discriminant ratio (FDR), ($r = 0.708$, $p = 0.010$). No significant correlations were rendered for overshoot.

2) Repeatability-Based Feedback: The Bhattacharyya distance (BD) yielded the strongest significant correlation value for throughput ($r = 0.686$, $p = 0.01$) and effective throughput ($r = 0.698$, $p = 0.01$). Path efficiency and overshoot demonstrated no significant correlations with any feature space metrics. The Kullback-Leibler divergence (KLD) was most correlated with average speed ($r = 0.788$, $p < 0.001$) while the intra-inter fraction (IIF) was most correlated with stopping distance ($r = -0.61$, $p = 0.04$).

3) Separability-Based Feedback: The repeatability index (RI) and mean semi-principal axes (MSA) were the only two feature space metrics to display significant correlations with the online metrics. The repeatability index demonstrated a significant correlation with throughput ($r = 0.585$, $p = 0.05$) while the mean semi-principal axes exhibited significant correlations with throughput ($r = -0.652$, $p = 0.02$), effective throughput ($r = -0.708$, $p = 0.01$), average speed ($r = -0.651$, $p = 0.02$), and stopping distance ($r = -0.592$, $p = 0.04$).

4) Functional-Based Feedback With Constant Velocity Control: The separability index (SI) produced the strongest significant correlation with throughput ($r = 0.685$, $p = 0.01$) and effective throughput ($r = 0.674$, $p = 0.02$). The mean within-trial separability index (mwSI) was most correlated with path efficiency ($r = 0.633$, $p = 0.03$) while compactness (C) was most correlated with average speed ($r = -0.605$, $p = 0.04$). No feature space metrics showed significant correlations with stopping distance.

5) Functional-Based Feedback With Proportional Control: Rescaled compactness (rC) produced the strongest significant correlation value for throughput ($r = -0.545$, $p = 0.01$). Hellinger distance (HD), training time (TT), and rescaled compactness (rC) produced equally high significant correlations for effective throughput ($r = -0.545$, $p = 0.01$). The intra-inter fraction (IIF) had the highest correlation with path efficiency ($r = 0.746$, $p = 0.005$). The mean within-repetition repeatability index (mwRI) was most correlated with overshoot ($r = 0.651$, $p = 0.02$) while the rescaled compactness (rC) measure was most correlated with average speed ($r = -0.515$, $p = 0.02$). Volume of overlap region (VOR) and rescaled compactness (rC) both demonstrated equally high significant correlations ($r = 0.545$, $p = 0.01$) with stopping distance.

D. Predictive Modeling

A summary of the predictive performance using the VE_{cv} metric is provided in Table II. Because a leave-one-subject-out (LOSO) cross-validation strategy was implemented to evaluate model performance, the model parameters and coefficients described in the following section differed slightly at each cross-validation fold. Therefore, rather than reporting the individual equations at each fold, the average relative coefficient weightings of the predictors across all folds are indicated in the text below. Positive and negative coefficient weightings indicate a direct and indirect relationship, respectively, with the response variable. Overall, the functional-based feedback with proportional control (FBF_{pc}) demonstrated the highest average VE_{cv} across all online performance metrics.

1) Throughput: The best predictive performance of the throughput response variable was observed when using SBF training metrics. The corresponding two- and three-predictor models outperformed those across the other feedback modalities. The VE_{cv} measured 82.9% for the three-predictor model, which is a 28% increase in performance compared to the corresponding model based on SGT training metrics and a 11.5% increase compared to the next best model using RBF training metrics. This top-performing model selected mean semi-principal axes (MSA), repeatability index (RI), and feature efficiency (FE) as its predictors. The average relative coefficient weighting for each predictor was 10.4%, 39.8%, and -49.8% for FE, RI, and MSA, respectively. The variability metrics MSA and RI collectively possessed about 90% of the relative coefficient weightings.

2) Effective Throughput: The SBF training metrics also held the greatest predictive power for effective throughput. The VE_{cv} measured 89.0% for the three-predictor model, which is a 36% increase in predictive performance compared to the

TABLE II

PREDICTIVE ACCURACY SUMMARY USING VE_{cv} VALUES ACROSS THE DIFFERENT FEEDBACK MODALITIES FOR EACH RESPONSE VARIABLE (RV): THROUGHPUT (TP), EFFECTIVE THROUGHPUT (eTP), PATH EFFICIENCY (PE), OVERSHOOT (OS), AVERAGE SPEED (AS), AND STOPPING DISTANCE (SD). CLASSIFICATION ACCURACY (CA) WAS SHOWN TO BE A POOR PREDICTOR OF USABILITY REGARDLESS OF THE FEEDBACK CONDITION. FBF_{pc} DEMONSTRATED THE HIGHEST AVERAGE VE_{cv} ACROSS ALL ONLINE PERFORMANCE METRICS FOR THE 1-, 2-, AND 3-PREDICTOR MODELS

RV	Model	SGT	RBF	SBF	FBF _{cv}	FBF _{pc}
TP	CA	-50.13	-13.39	-55.94	-32.84	-55.77
	1 Predictor	17.41	25.59	19.01	22.31	17.51
	2 Predictors	46.31	57.19	79.89	68.88	54.07
	3 Predictors	64.67	74.31	82.85	64.47	65.52
eTP	CA	-49.52	-11.74	-56.18	-32.45	-54.53
	1 Predictor	15.54	27.49	31.01	19.71	16.66
	2 Predictors	47.42	55.17	86.38	65.72	48.01
	3 Predictors	65.63	77.57	89.04	74.61	65.29
PE	CA	-14.95	-29.52	-47.39	-18.03	-84.95
	1 Predictor	15.02	42.95	-4.40	12.99	66.42
	2 Predictors	41.16	67.72	7.15	39.49	77.88
	3 Predictors	64.39	71.36	21.81	71.17	91.18
OS	CA	-33.35	-31.18	-37.01	-36.61	-54.53
	1 Predictor	-3.04	-3.26	-11.95	13.69	38.48
	2 Predictors	-3.25	41.63	-5.60	78.59	82.23
	3 Predictors	50.55	63.69	5.43	91.52	87.75
AS	CA	-29.68	-9.86	-50.59	-43.99	-58.51
	1 Predictor	20.29	36.51	19.61	13.86	10.64
	2 Predictors	43.72	69.13	72.47	37.05	53.03
	3 Predictors	59.77	75.09	78.41	79.88	71.96
SD	CA	-3.53	-176.4	-43.52	-24.59	-91.97
	1 Predictor	18.66	28.79	3.71	5.76	45.56
	2 Predictors	38.96	46.07	61.36	34.27	52.76
	3 Predictors	52.59	60.94	70.97	50.73	63.75
Avg.	CA	-30.19	-45.35	-48.44	-31.42	-66.71
	1 Predictor	13.98	26.35	9.50	14.72	32.54
	2 Predictors	35.72	56.15	50.27	54.00	61.33
	3 Predictors	59.60	70.49	58.08	72.06	74.24

corresponding model based on SGT training metrics and a 15% increase compared to the next best model using RBF training metrics. The top-performing model selected the same predictors as the model for throughput with near equal relative coefficient weightings. The average relative coefficient weighting for each predictor was 8.7%, 38.3%, and -53.0% for FE, RI, and MSA, respectively. Over 90% of the relative coefficient weightings were attributed to the same two variability metrics as in the model for throughput, MSA and RI.

3) *Path Efficiency*: The FBF_{pc} protocol led to the highest predictability for path efficiency. The VE_{cv} measured 91.2% for the three-predictor model, which is a 42% increase in predictive performance compared to the corresponding model based on SGT training metrics and a 28% increase compared to the next best model using RBF training metrics. The selected predictors for this model were intra-inter fraction (IIF), collective entropy (CE), and purity (PU). The average relative coefficient weighting for each predictor was 10.9%, 42.9%, and -46.2% for IIF, PU, and CE, respectively. The complexity metrics CE and PU possessed 89% of the relative coefficient weightings.

4) *Overshoot*: The best predictive performance of overshoot was generated during the FBF_{cv} training session. The VE_{cv} measured 91.5% for the three-predictor model, equaling an 81% increase in predictive performance compared to the corresponding model based on SGT training metrics and a 4% increase compared to the next best model using FBF_{pc} training metrics. The selected predictors for this three-predictor model were the repeatability index (RI), classification accuracy (CA), and the class discriminability metric (CDM). The average relative coefficient weighting for each predictor was 53.5%, -24.8%, and -21.7% for RI, CA, and CDM, respectively.

5) *Average Speed*: Average speed demonstrated the most accurate predictability by leveraging the FBF_{cv} training metrics. The VE_{cv} measured 79.9% for the three-predictor model, which is a 34% increase in predictive performance compared to the corresponding model based on SGT training metrics and a 2% increase compared to the next best model using SBF training metrics. The best-selected predictors for this model were mean semi-principal axes (MSA), intra-class fraction (ICF), and compactness (C). The average relative coefficient weighting for each predictor was 20.3%, -24.7%, and 55.0% for MSA, ICF, and C, respectively.

6) *Stopping Distance*: The SBF training session yielded the greatest predictive performance for stopping distance. The VE_{cv} measured 71.0% for the three-predictor model, which is a 35% increase in predictive performance compared to the corresponding model based on SGT training metrics and an 11% increase compared to the next best model using FBF_{pc} training metrics. The best-selected predictors for this model were feature efficiency (FE), active classification accuracy (ACA), and the class discriminability metric (CDM). The average relative coefficient weighting for each predictor was 33.1%, -39.1%, and -27.8% for FE, ACA, and CDM, respectively.

IV. DISCUSSION

This study investigated the effect of presenting feedback during user training on both the online performance of a pattern recognition-based myoelectric control system and on the predictability of that performance. Providing users with knowledge of the system's behavior during the training process may lead to 1) contractions that better represent those seen during real-time use and 2) an improved understanding of what is necessary for successful pattern recognition control. A more predictable control system could lead to more reliable use and, in turn, a positive impact on the adoption of myoelectrically-controlled devices.

A. Effect of Feedback on Feature Space

The type of feedback presented during training had a significant impact on feature space characteristics. The FBF modalities often yielded feature space metrics that were significantly worse than the other feedback types. These results were expected because users were required to correct for unintended movements and misclassifications. Additionally, the contractions during FBF were dynamic as users ramped up to initiate cursor movement and down to slow the cursor to a

halt on the target. Previous work has demonstrated that offline classification error increases when these transient portions of the EMG signal are incorporated into the training dataset [44]. The dynamic contractions associated with the FBF protocols also explain why the FBF training data produced higher MSA and lower SI, as shown in Figures 3d and 3c. MSA is a measure of hyperellipsoid size, so as variability is introduced into the training data, MSA increases [16]. Changes in MSA also contribute to the separability of the training data [16]. Therefore, the high MSA observed by the FBF training data is a likely cause for low separability metrics.

B. Effect of Feedback on Online Performance

The visual feedback modalities tested in this study typically did not lead to any significant improvements in online performance as compared to the screen guided training (SGT) approach.

Other studies implementing similar feedback conditions have demonstrated mixed results. Kristoffersen *et al.* and Roche *et al.* both employed feedback similar to our repeatability-based feedback [24], [40]. Kristoffersen *et al.* tested a classification scheme using three levels of feedback: no external feedback, visual feedback and visual feedback with coaching. They found that user training with normally-limbed subjects resulted in improved performance from pre-test to post-test, but there was no interaction effect with regards to the feedback group [24]. However, Roche *et al.* did find in their case study that a structured rehabilitation protocol involving imitation, repetition and reinforcement with visual feedback led to an improvement in a transradial amputee's Southampton Hand Assessment Procedure (SHAP) from 58 to 71 in a simultaneous control scheme.

Fang *et al.* implemented a clustering-based feedback similar to our separability-based feedback [41]. They found that it led to significantly higher classification accuracy than the no feedback case and the classifier label feedback case. One thing to note is that they did not start collecting EMG signals until five seconds after the cue signal to allow users time to respond and position the cursor. Given that we found in our study that classification accuracy is a poor predictor of usability, regardless of the feedback condition, it is unclear how well their classifier would respond during real-time use.

Fitts' Law style testing is a popular approach for evaluating myoelectric control performance [7], [55], [56], [63]; however, to the best of our knowledge, there are no published results implementing a Fitts' Law targeting paradigm as part of the training phase. While the Fitts' Law target acquisition training protocol implemented in our functional-based feedback (FBF) approach yielded "poor" feature space metrics, FBF_{cv} produced the highest mean throughput and effective throughput while FBF_{pc} produced the highest median path efficiency, the lowest median overshoot, and lowest mean stopping distance (although no significant differences were observed compared to SGT). These results support previous findings that, despite decreasing offline classification accuracy, introducing variability into the training data can lead to increases in online performance [44], [64]. Based on the findings described in

Section IV-A, this same idea may be true for other feature space metrics as well.

Although the results were not significant, it makes sense that the FBF strategy would lead to improved online performance. When training with FBF, users were exposed to an environment similar to the one used during testing. This may have enabled users to become familiar with the dynamics of their movements prior to testing and allowed them to build an internal model of the system's behavior [22]. In particular, FBF_{pc} allowed users to practice modulating their contraction intensity to control the speed of the cursor.

Therefore, it is not unexpected that online performance tended to improve as the training environment approached the testing conditions.

It is interesting to note here that the FBF_{pc} led to the lowest throughput, effective throughput, and average speed. We observed that, because the user had proportional control during training, their effort level was higher than what it was for the other feedback modalities. Because their training contractions were high amplitude, it was more difficult for them to surpass that contraction intensity during the testing phase to initiate faster cursor movement. Therefore, we expect that the low usability scores for these metrics are more the result of the proportional control parameter settings than the result of the feedback on user behavior. This effect could be compensated for in future studies by using a proportional control gain adjustment [65].

C. Effect of Session on Performance

Several studies have reported that classification accuracy (CA) improves with user practice [12], [16], [35]. However, we found no significant differences in CA or any of the other feature space metrics across the five sessions. Improvements in these metrics may have been limited by the use of an initial familiarity session or inhibited because users were exposed to different forms of feedback during every session. As described in Section IV-A, the type of feedback presented during training significantly impacted training data characteristics. Therefore, techniques learned during one training session may not have translated to another feedback condition in a different session.

Online myoelectric control also has been shown to improve with additional training. A recent study demonstrated that following a 10-day training protocol, a perceived performance plateau was reported no earlier than the fifth training session [17]. Supporting the results of Tabor *et al.* [17], Figure 5 demonstrates a visible increasing trend in performance for throughput, effective throughput, and average speed. These results suggest that users become more experienced and confident with the testing task in each subsequent session independent of the observed feedback.

D. Effect of Feedback on Predicting Usability

Although the effect of feedback on myoelectric control performance has been explored by many researchers [15], [16], [23], [24], [35], the effect of feedback on the predictability of

classification-based myoelectric control is a relatively undocumented area of research. The ability to predict user performance could have significant implications for the adoption of myoelectrically controlled devices. The reasons for low user acceptance of upper-limb myoelectric prostheses have been a topic of interest for decades [1], [66], [67]. Erroneous motions and limitations in the dexterity of control are among some of the common attributing factors for abandonment [2], [6]. If users knew that the outcome of a training protocol might result in poor practical use later on, they could retrain the system immediately to avoid unnecessary frustration during activities of daily living.

The results of this study suggest that certain user-in-the-loop training schemes can influence the characteristics of the training data to be more indicative of future real-time use without any significant degradation in performance. For each response variable, the training data that led to the best predictive performance was generated from one of the four feedback modalities. Although a single best feedback would have been preferable, results showed that the type of feedback strategy that led to the best predictive performance differed depending on the online metric being predicted.

The SBF training scheme produced feature space metrics that were best able to predict throughput, effective throughput, and stopping distance. The remaining three online performance metrics were best predicted using the FBF training schemes. The FBF_{cv} training provided the most informative data for predicting overshoot and average speed whereas the FBF_{pc} condition influenced the training metrics in a way that best supported predictions of path efficiency. Because the FBF training schemes parallel the testing environment, the user was aware of the system's dynamics prior to the evaluation phase. If the user generated poor path efficiency or overshoot during training, it would be expected that the same result would occur during testing and that the training data would somehow reflect this behavior.

The largest variability in VE_{cv} was observed for the response variable overshoot. For the three-predictor overshoot models, SBF produced the lowest predictive accuracy at 5.43%, and both FBF conditions produced the highest predictive accuracy at 87.75% and 91.52% for FBF_{pc} and FBF_{cv}, respectively. The ability for users to practice acquiring targets during FBF training may have promoted the formation of task-specific internal models that better reflected the testing environment [68]. If the internal models developed during SGT, RBF, and SBF were not representative of the testing task, then user behavior could have been unpredictable during testing, resulting in the observed decreases in predictive performance.

While absolute myoelectric control performance is an important aspect of user satisfaction, having consistent and predictable performance is also important. The results from this study show that training with certain feedback can provide a system that encompasses both.

Under the conditions of this study, the FBF_{pc} protocol resulted in the most robust system, combining desirable

attributes of predictability and usability. FBF_{pc} yielded the highest average VE_{cv} score across the six online metrics for the one-, two-, and three-predictor models while also obtaining the highest online performance for three out of the six online metrics. It is possible that the remaining three online metrics were confounded by the proportional control parameters and may not have reflected the true user behaviors in response to the feedback.

Given that the functional-based feedback conditions generated training data that best described the testing performance, however, developing training protocols with similar conditions as those experienced during real-time use may provide users with a more informed expectation of the system's output.

E. Limitations and Future Work

In this study, data from twelve participants were collected. This relatively small sample size did not allow for separate training, validation and test sets. Instead, a leave-one-subject-out cross validation was implemented for model evaluation. This approach is commonly used in the literature, but may not produce the most generalizable results. Therefore, the reported predictive performance could be an inflated view of the true model behavior. However, given that the sample size remained constant across all feedback conditions and models, the general impact of the feedback on the feature space metrics and the ability to use those metrics as predictors of usability is likely a stable result. Future work should aim to validate these results by expanding the subject pool to include many more samples.

Carry-over and sequence effects also may have confounded the online performance results. These effects are the primary disadvantage of a within-subject experimental design. A carry-over effect results when the effect of being tested in one experimental condition influences a participant's behavior in subsequent conditions [69]. Learning is a common source of carry-over effect and was observed in this study.

Sequence effects are characterized by potential interactions among experimental conditions depending on the order in which the conditions are presented [69]. If experiencing a particular condition before another significantly influences performance on the next condition, sequence effects may be present. Counterbalancing is one method that attempts to minimize these effects by ensuring that the effects fall equally on all conditions [69].

To completely eliminate carry-over and sequence effects, future work could look at implementing a between-subject experimental design. This design approach could also allow an investigation into the learning rates of users for each feedback by incorporating multiple sequential sessions. Therefore, feedback would be a between subject variable and session would be a within-subject variable.

F. Conclusion

The results of this study suggest that the use of augmented visual feedback during user training can significantly influence training data characteristics and, in many cases, amplify the

predictability of a myoelectric control system. By incorporating the user in the control loop, users have the ability to adjust and regulate their muscle contractions based on the decisions made by the classifier. This may, in turn, make the training data more representative of real-time use, and lead to better predictability.

REFERENCES

- [1] E. A. Biddiss and T. T. Chau, "Upper limb prosthesis use and abandonment: A survey of the last 25 years," *Prosthetics Orthotics Int.*, vol. 31, no. 3, pp. 236–257, 2007.
- [2] A. Chadwell, L. Kenney, S. Thies, A. Galpin, and J. Head, "The reality of myoelectric prostheses: Understanding what makes these devices difficult for some users to control," *Frontiers Neurobotics*, vol. 10, p. 7, Aug. 2016.
- [3] R. J. Beaulieu *et al.*, "Multi-position training improves robustness of pattern recognition and reduces limb-position effect in prosthetic control," *J. Prosthetics Orthot.*, vol. 29, no. 2, pp. 54–62, Apr. 2017.
- [4] A. Boschmann and M. Platzner, "Reducing the limb position effect in pattern recognition based myoelectric control using a high density electrode array," in *Proc. ISSNIP Biosignals Biorobot. Conf. (BRC)*, Feb. 2013, pp. 1–5.
- [5] A. Fougner, E. Scheme, A. D. C. Chan, K. Englehart, and Ø. Stavdahl, "Resolving the limb position effect in myoelectric pattern recognition," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 19, no. 6, pp. 644–651, Dec. 2011.
- [6] E. Scheme and K. Englehart, "Electromyogram pattern recognition for control of powered upper-limb prostheses: State of the art and challenges for clinical use," *J. Rehabil. Res. Dev.*, vol. 48, no. 6, pp. 643–660, Jan. 2011.
- [7] E. J. Scheme and K. B. Englehart, "Validation of a selective ensemble-based classification scheme for myoelectric control using a three-dimensional Fitts' law test," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 4, pp. 616–623, Jul. 2013.
- [8] E. Campbell, A. Phinyomark, and E. Scheme, "Current trends and confounding factors in myoelectric control: Limb position and contraction intensity," *Sensors*, vol. 20, no. 6, p. 1613, Mar. 2020.
- [9] L. Hargrove, K. Englehart, and B. Hudgins, "The effect of electrode displacements on pattern recognition based myoelectric control," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2006, pp. 2203–2206.
- [10] L. Hargrove, K. Englehart, and B. Hudgins, "A training strategy to reduce classification degradation due to electrode displacements in pattern recognition based myoelectric control," *Biomed. Signal Process. Control*, vol. 3, no. 2, pp. 175–180, 2008.
- [11] J. W. Sensinger, B. A. Lock, and T. A. Kuiken, "Adaptive pattern recognition of myoelectric signals: Exploration of conceptual framework and practical algorithms," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 17, no. 3, pp. 270–278, Jun. 2009.
- [12] J. He, D. Zhang, N. Jiang, X. Sheng, D. Farina, and X. Zhu, "User adaptation in long-term, open-loop myoelectric training: Implications for EMG pattern recognition in prosthesis control," *J. Neural Eng.*, vol. 12, no. 4, Aug. 2015, Art. no. 046005.
- [13] H. Zhang, Y. Zhao, F. Yao, L. Xu, P. Shang, and G. Li, "An adaptation strategy of using LDA classifier for EMG pattern recognition," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2013, pp. 4267–4270.
- [14] S. M. Radhakrishnan, S. N. Baker, and A. Jackson, "Learning a novel myoelectric-controlled interface task," *J. Neurophysiol.*, vol. 100, no. 4, pp. 2397–2408, Oct. 2008.
- [15] A. Krasoulis, S. Vijayakumar, and K. Nazarpour, "Effect of user practice on prosthetic finger control with an intuitive myoelectric decoder," *Frontiers Neurosci.*, vol. 13, pp. 1–16, Sep. 2019.
- [16] N. E. Bunderson and T. A. Kuiken, "Quantification of feature space changes with experience during electromyogram pattern recognition control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 3, pp. 239–246, May 2012.
- [17] A. Tabor, S. Bateman, and E. Scheme, "Evaluation of myoelectric control learning using multi-session game-based training," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 9, pp. 1680–1689, Sep. 2018.
- [18] J. M. Hahne, S. Dähne, H.-J. Hwang, K.-R. Müller, and L. C. PARRA, "Concurrent adaptation of human and machine improves simultaneous and proportional myoelectric control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 23, no. 4, pp. 618–627, Jul. 2015.
- [19] C. Igual, L. A. Pardo, J. M. Hahne, and J. Igual, "Myoelectric control for upper limb prostheses," *Electronics*, vol. 8, no. 11, p. 1244, Nov. 2019.
- [20] M. Couraud, D. Cattaert, F. Paclét, P. Y. Oudeyer, and A. de Ruyg, "Model and experiments to optimize co-adaptation in a simplified myoelectric control system," *J. Neural Eng.*, vol. 15, no. 2, Apr. 2018, Art. no. 026006.
- [21] J. S. Müller, C. Vidaurre, M. Schreuder, F. C. Meinecke, P. von Büna, and K.-R. Müller, "A mathematical model for the two-learners problem," *J. Neural Eng.*, vol. 14, no. 3, Jun. 2017, Art. no. 036005.
- [22] A. W. Shehata, L. F. Engels, M. Controzzi, C. Cipriani, E. J. Scheme, and J. W. Sensinger, "Improving internal model strength and performance of prosthetic hands using augmented feedback," *J. NeuroEng. Rehabil.*, vol. 15, no. 1, pp. 1–12, Jul. 2018.
- [23] A. M. Simon, B. A. Lock, and K. A. Stubblefield, "Patient training for functional use of pattern recognition–controlled prostheses," *J. Prosthetics Orthotics*, vol. 24, no. 2, pp. 56–64, 2012.
- [24] M. B. Kristoffersen, A. W. Franzke, C. K. van der Sluis, A. Murgia, and R. M. Bongers, "The effect of feedback during training sessions on learning pattern-recognition-based prosthesis control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 10, pp. 2087–2096, Oct. 2019.
- [25] F. V. G. Tenore, A. Ramos, A. Fahmy, S. Acharya, R. Etienne-Cummings, and N. V. Thakor, "Decoding of individuated finger movements using surface electromyography," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 5, pp. 1427–1434, May 2009.
- [26] M. Markovic *et al.*, "The clinical relevance of advanced artificial feedback in the control of a multi-functional myoelectric prosthesis," *J. Neuroeng. Rehabil.*, vol. 15, pp. 1–15, Mar. 2018.
- [27] A. S. R. Parker, A. L. Edwards, and P. M. Pilarski, "Exploring the impact of machine-learned predictions on feedback from an artificial limb," in *Proc. IEEE 16th Int. Conf. Rehabil. Robot. (ICORR)*, Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers, Jun. 2019, pp. 1239–1246.
- [28] J. González and W. Yu, "Multichannel audio aided dynamical perception for prosthetic hand biofeedback," in *Proc. IEEE Int. Conf. Rehabil. Robot.*, Jun. 2009, pp. 240–245.
- [29] A. E. Davis and R. G. Lee, "EMG biofeedback in patients with motor disorders: An aid for co-ordinating activity in antagonistic muscle groups," *Can. J. Neurol. Sci.*, vol. 7, no. 3, pp. 199–206, Aug. 1980.
- [30] R. E. Schleenbaker and A. G. Mainous, "Electromyographic biofeedback for neuromuscular reeducation in the hemiplegic stroke patient: A meta-analysis," *Arch. Phys. Med. Rehabil.*, vol. 74, no. 12, pp. 1301–1304, 1993.
- [31] B. S. Brucker and N. V. Buylaeva, "Biofeedback effect on electromyography responses in patients with spinal cord injury," *Arch. Phys. Med. Rehabil.*, vol. 77, no. 2, pp. 133–137, Feb. 1996.
- [32] S. Dosen, M. Markovic, K. Somer, B. Graimann, and D. Farina, "EMG biofeedback for online predictive control of grasping force in a myoelectric prosthesis," *J. NeuroEng. Rehabil.*, vol. 12, no. 1, pp. 1–13, Dec. 2015.
- [33] M. I. P. Lourenção, L. R. Battistella, C. M. M. de Brito, G. R. Tsukimoto, and M. H. Miyazaki, "Effect of biofeedback accompanying occupational therapy and functional electrical stimulation in hemiplegic patients," *Int. J. Rehabil. Res.*, vol. 31, no. 1, pp. 33–41, 2008.
- [34] F. Tamburella *et al.*, "Influences of the biofeedback content on robotic post-stroke gait rehabilitation: Electromyographic vs joint torque biofeedback," *J. NeuroEng. Rehabil.*, vol. 16, no. 1, Dec. 2019, Art. no. 95.
- [35] M. A. Powell, R. R. Kaliki, and N. V. Thakor, "User training for pattern recognition-based myoelectric prostheses: Improving phantom limb movement consistency and distinguishability," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 3, pp. 522–532, May 2014.
- [36] M. Ison and P. Artemiadis, "Proportional myoelectric control of robots: Muscle synergy development drives performance enhancement, retention, and generalization," *IEEE Trans. Robot.*, vol. 31, no. 2, pp. 259–268, Apr. 2015.
- [37] M. Ison, I. Vujaklija, B. Whitsell, D. Farina, and P. Artemiadis, "High-density electromyography and motor skill learning for robust long-term control of a 7-DoF robot arm," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 4, pp. 424–433, Apr. 2016.
- [38] J. M. Hahne *et al.*, "Simultaneous and proportional control of 2D wrist movements with myoelectric signals," in *Proc. IEEE Int. Workshop Mach. Learn. Signal Process. (MLSP)*, Sep. 2012, pp. 1–6.
- [39] T. Pistohl, C. Cipriani, A. Jackson, and K. Nazarpour, "Abstract and proportional myoelectric control for multi-fingered hand prostheses," *Ann. Biomed. Eng.*, vol. 41, no. 12, pp. 2687–2698, Dec. 2013.

- [40] A. D. Roche *et al.*, "A structured rehabilitation protocol for improved multifunctional prosthetic control: A case study," *J. Vis. Exp.*, vol. 2015, no. 105, pp. 1–8, Nov. 2015.
- [41] Y. Fang, D. Zhou, S. Member, K. Li, and S. Member, "Classifier-feedback-based user training," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 11, pp. 2575–2583, Nov. 2017.
- [42] J. Vance, G. Wulf, T. Tollner, N. Mcnevin, and J. Mercer, "EMG activity as a function of the performer's focus of attention," *J. Motor Behav.*, vol. 36, no. 4, pp. 450–459, 2004.
- [43] B. A. Lock, K. Englehart, and B. Hudgins, "Real-time myoelectric control in a virtual environment to relate usability vs. accuracy," in *Proc. MyoElectr. Control. Prosthetics Symp.*, 2005, pp. 1–4.
- [44] L. Hargrove, Y. Losier, B. Lock, K. Englehart, and B. Hudgins, "A real-time pattern recognition based myoelectric control usability study implemented in a virtual environment," in *Proc. 29th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2007, pp. 4842–4845.
- [45] N. Jiang, I. Vujaklija, H. Rehbaum, B. Graimann, and D. Farina, "Is accurate mapping of EMG signals on kinematics needed for precise online myoelectric control?" *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 3, pp. 549–558, May 2014.
- [46] N. Nilsson, B. Håkansson, and M. Ortiz-Catalan, "Classification complexity in myoelectric pattern recognition," *J. NeuroEng. Rehabil.*, vol. 14, no. 1, Jul. 2017, Art. no. 68.
- [47] J. L. Nawfel, K. B. Englehart, and E. J. Scheme, "A multi-variate approach to predicting myoelectric control usability," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1312–1327, 2021.
- [48] E. Scheme and K. Englehart, "A flexible user interface for rapid prototyping of advanced real-time myoelectric control schemes," in *Proc. MyoElectr. Control. Prosthetics Symp.*, 2008, pp. 1–4.
- [49] B. Hudgins, P. Parker, and R. N. Scott, "A new strategy for multifunction myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 40, no. 1, pp. 82–94, Jan. 1993.
- [50] A. Bhattacharyya, "On a measure of divergence between two multinomial populations," *Indian J. Stat.*, vol. 7, no. 4, pp. 401–406, 1946.
- [51] T. K. Ho and M. Basu, "Complexity measures of supervised classification problems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 3, pp. 289–300, Mar. 2002.
- [52] A. F. Kohn, L. G. Nakano, and M. O. E. Silva, "A class discriminability measure based on feature space partitioning," *Pattern Recognit.*, vol. 29, no. 5, pp. 873–887, 1996.
- [53] N. E. Krausz, B. H. Hu, and L. J. Hargrove, "Subject- and environment-based sensor variability for wearable lower-limb assistive devices," *Sensors*, vol. 19, no. 22, p. 4887, Nov. 2019.
- [54] S. Singh, "PRISM—A novel framework for pattern recognition," *Pattern Anal. Appl.*, vol. 6, pp. 134–149, Jun. 2003.
- [55] J. Park, W. Bae, H. Kim, and S. Park, "EMG—Force correlation considering Fitts' law," in *Proc. IEEE Int. Conf. Multisensor Fusion Integr. Intell. Syst.*, Aug. 2008, pp. 644–649.
- [56] J. Gusman, E. Mastinu, and M. Ortiz-Catalan, "Evaluation of computer-based target achievement tests for myoelectric control," *IEEE J. Transl. Eng. Health Med.*, vol. 5, pp. 1–10, 2017.
- [57] P. M. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement," *J. Exp. Psychol.*, vol. 47, no. 6, pp. 381–391, Jun. 1954.
- [58] M. R. Williams and R. F. Kirsch, "Evaluation of head orientation and neck muscle EMG signals as command inputs to a human-computer interface for individuals with high tetraplegia," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 16, no. 5, pp. 485–496, Oct. 2008.
- [59] H. Khamis, "Measures of association: How to choose?" *J. Diagnostic Med. Sonogr.*, vol. 24, no. 3, pp. 155–162, May 2008.
- [60] S. Parvande, H.-W. Yeh, M. P. Paulus, and B. A. McKinney, "Consensus features nested cross-validation," *Bioinformatics*, vol. 36, no. 10, pp. 3093–3098, May 2020.
- [61] J. Li, "Assessing spatial predictive models in the environmental sciences: Accuracy measures, data variation and variance explained," *Environ. Model. Softw.*, vol. 80, pp. 1–8, Jun. 2016.
- [62] J. Li, "Assessing the accuracy of predictive models for numerical data: Not r nor r^2 , why not? Then what?" *PLoS ONE*, vol. 12, no. 8, Aug. 2017, Art. no. e0183250.
- [63] A. Ameri, M. A. Akhaee, E. Scheme, and K. Englehart, "Real-time, simultaneous myoelectric control using a convolutional neural network," *PLoS ONE*, vol. 13, no. 9, pp. 1–13, 2018.
- [64] L. J. Hargrove, E. J. Scheme, K. B. Englehart, and B. S. Hudgins, "Multiple binary classifications via linear discriminant analysis for improved controllability of a powered prosthesis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 18, no. 1, pp. 49–57, Feb. 2010.
- [65] L. Resnik, H. H. Huang, A. Winslow, D. L. Crouch, F. Zhang, and N. Wolk, "Evaluation of EMG pattern recognition for upper limb prosthesis control: A case study in comparison with direct myoelectric control," *J. NeuroEng. Rehabil.*, vol. 15, no. 1, pp. 1–13, Dec. 2018.
- [66] D. J. Atkins, D. C. Y. Heard, and W. H. Donovan, "Epidemiologic overview of individuals with upper-limb loss and their reported research priorities," *J. Prosthetics Orthotics*, vol. 8, no. 1, pp. 2–11, 1996.
- [67] S. Millstein, H. Heger, and G. Hunter, "A review of the failures in use of the below elbow myoelectric prosthesis," *Orthot Prosthetics*, vol. 36, no. 2, pp. 29–34, 1982.
- [68] C. Tong and J. R. Flanagan, "Task-specific internal models for kinematic transformations," *J. Neurophysiol.*, vol. 90, no. 2, pp. 578–585, Aug. 2003.
- [69] E. F. Corriero, *The SAGE Encyclopedia of Communication Research Methods*, M. Allen, Ed. Thousand Oaks, CA, USA: Sage, 2017.



DSPACE

<https://dspace.org/>

The Influence of Training With Visual Biofeedback on the Predictability of Myoelectric Control Usability

Nawfel, Jena L.; Englehart, Kevin B.; Scheme, Erik J.

2022

Institute of Electrical and Electronics Engineers

<https://unbscholar.lib.unb.ca/handle/1882/37226>