2021 10th International IEEE/EMBS Conference on Neural Engineering (NER)
Virtual Conference, May 4-6, 2021

Classifiers and Adaptable Features Improve Myoelectric Command Accuracy in Trained Users*

Sarah M. O'Meara, *Student Member, IEEE,* Stephen K. Robinson, and Sanjay S. Joshi, *Senior Member, IEEE*

2021 10th International IEEE/EMBS Conference on Neural Engineering (NER) | 978-1-7281-4337-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/NER49283.2021.9441256

Abstract— **A common challenge in myoelectric control is to create reliable human-machine interfaces. The aim of our study is to apply classifiers and adaptable features to electromyography (EMG) signals obtained from subjects who are trained to perform cursor-control tasks. We have developed a robust surface EMG command system that relies on training the human rather than training a classifier, thereby enabling the human to correct for signal changes. The purpose of the current study was to understand whether adding an adaptive timing feature and classifiers to our EMG interface could improve the performance of trained human subjects. Forty-eight subjects participated in the experiment, where they learned four different commands to control a cursor to select fixed targets on a computer screen, where each command was composed of a combination of short and long muscle contractions. A Command Accuracy Test assessed subject proficiency at producing commands when prompted. The command classification accuracy was calculated for a control condition and two conditions that reflected possible adaptive features: the timing between EMG signal inputs and a subject-specific classifier. The overall results showed significant improvements in command classification accuracy for both adaptive components (*$p <$* 0.0001) compared to the control. However, some initially high performing subjects did not receive as much benefit. These results suggest that customizing the sEMG command system for individual subjects could improve their performance. Future work should investigate the effect of customizing the system for performance and co-adaptation, as well as using the adaptive features as a training tool to further improve command accuracy and efficiency.**

I. Introduction

Myoelectric control, or the use of electromyographic (EMG) signals to convey a command, has been used in prosthetics and human-computer interfaces, with various strategies employed to create a reliable communication interface. However, EMG can suffer from signal non-stationarities due to subject fatigue, shifting electrodes, and postural changes [1]. Some researchers have mitigated signal non-stationarities by developing adaptive systems. These adaptive systems have been shown to improve pattern recognition classifier accuracy during increased signal noise [2], to augment performance with online co-adaptive learning [3][4], and to maintain repeatability over multiple sessions [5]. All of these research studies used surface EMG (sEMG) control with multiple sensors on the forearm, and most used pattern recognition [2] or regression-based methods [3][4].

Adaptive sEMG systems can focus on mitigating signal non-stationarities and/or improving overall system accuracy and performance. Benchabane et al. [5] developed a system for simultaneous control of prosthetic fingers using an adaptive threshold to select one or more finger functions and demonstrated low accuracy loss, despite multiple sessions (-4.8%) and subject fatigue (-2.4%). Similarly, Zhang et al. [2] tested their system with several noise levels to compare a Linear Discriminant Analysis (LDA) classifier with an adaptive LDA (ALDA) that periodically incorporated new training data to recalculate parameters online. The ALDA outperformed the LDA across conditions [2]. Hahne et al. [3] investigated different adaptation rates for a co-adaptation cursor-to-target task in a regression-based system to improve task performance. In a similar regression-based system, Yeung et al. [4] compared different methods of weighting new training data, tested several adaptation rates, and demonstrated improved performance compared to an unassisted user. Taken together, these selected studies demonstrate the promise of adaptive sEMG systems for novice users. Our work was motivated to investigate the benefits of adaptive systems for users after completing training and reaching a learning plateau to further enhance their performance.

Previously, we had developed a single-site sEMG cursor control method that allowed the user to select among four cursor commands, using a coded sequence of short and long muscle contractions. This method may be less sensitive to signal non-stationarities due to the emphasis on user training. The user only needs to cross a threshold to communicate and can adjust their own actions to make corrections. However, this approach placed the burden of learning on user training, and we subsequently investigated the effects of various training strategies on performance, workload, and trust [6]. At the end of the training period, subjects across training strategies achieved an average completion rate of 0.86 and command classification accuracy of 66.15% for a cursor-to-target task [6]. The purpose of the work presented here was to understand if adding an adaptive timing component and individually trained classifiers would improve the classification accuracy of trained subjects.

II. Methods

A. Experiment Design and Setup

In this study, we utilized the data collected in [6] to explore the possibility of integrating an adaptive feature or classifier to

* This material is based upon work supported by the National Science Foundation under Grant No. (1934792). This work was supported by a Space Technology Research Institutes grant from NASA's Space Technology Research Grant Program. Support also provided from the Link Foundation (Link Fellowship in Link Foundation Modeling, Simulation, and Training).

S. M. O'Meara, S. K. Robinson, and S. S. Joshi are with the Department of Mechanical and Aerospace Engineering, University of California Davis, Davis, CA, 95616 (e-mail: someara@ucdavis.edu, stephen.k.robinson@ucdavis.edu, maejoshi@ucdavis.edu).

our sEMG control method. (The detailed experiment design and protocol can be found in [6].) The forty-eight subjects that completed the experiment ranged in age from 18 to 23 years with two left-hand dominant subjects and an equal participation of men and women. Subjects were able-bodied university students who reported no history of neuromuscular disorders or myoelectric control experience. The University of California Davis Institutional Review Board approved the experimental protocol.

The experimental setup included two electrodes (ConMed 1620 Ag/AgCl center snap) placed approximately 2.5 cm apart near the extensor digitorum proximal attachment on the dominant side. The reference electrode was fixed close to the lateral epicondyle of the humerus. The signal acquisition followed [7] and was processed as in [8], where the signal was sampled at 4096 Hz for 256-sample windows with a Butterworth filter (4th order, bandpass at 10 Hz and 500 Hz). The rms value for the window was normalized by each subject's calibration value and put in a moving average filter (length 0.5 s) for a processed signal update rate of 16 Hz.

B. Cursor Control and Training Methods

The sEMG method consisted of four cursor-control commands—up, down, left, right—that the subjects selected through a serial pattern akin to Morse code. The commands consisted of two consecutive signal inputs with a rest between each input; if the rest exceeded the timeout (0.5 s), the pattern reset (see Fig. 1A). An input started when the processed signal exceeded a threshold, l_1, and ended when the processed signal returned to below the threshold. Inputs less than 0.5 s were categorized as "short" and inputs greater than 0.5 s were "long". The combined pair of short and long inputs (a.k.a. muscle contractions) defined a command; for example, two short inputs selected the "up" command (see Fig. 1B). A third input moved the cursor forward in the selected direction while the processed signal remained above the threshold.

During the training portion of the experiment, the subjects learned to control a cursor on a computer screen to select targets in a Fitts's law style task. There were 120 training trials and 40 evaluation trials, where a trial was a single cursor-to-target attempt. The training strategy was in effect during the training trials and removed during evaluation. Subjects were assigned to one of four groups that determined the training strategy: Repetition, Concurrent Feedback, Terminal Feedback, and Adaptive Threshold.

- The Repetition group learned only by repeatedly performing the cursor-to-target task.
- The Concurrent Feedback group received an additional visual feedback element, where the cursor body changed color when the processed signal exceeded the threshold shown in Fig. 1A.
- The Terminal Feedback group's additional visual feedback element indicated the selected command at the end of Input 2 (see Fig. 1A).
- The Adaptive Threshold group saw the same interface as the Repetition group, but the threshold varied trial-by-trial (l_1 = 0.10, 0.15, 0.20, 0.25, 0.30). The threshold was set to 0.20 for other groups (threshold shown in Fig. 1A).

C. Analysis

To assess the subjects' proficiencies in selecting commands, a Command Accuracy Test was administered before training, after training (Test 2), and after evaluation (Test 3). Subjects were prompted to produce each command 5 times for a total of 20 attempts per test. Subjects were only allowed one try within an attempt. Command classification accuracy was calculated as the number of successfully selected commands out of the total for each test. The data from the Command Accuracy Tests 2 and 3 were combined and used for the analysis presented here. From the previous study, there were no significant differences after the training period between Tests 2 and 3 [6]. A larger dataset would have been ideal, but this analysis was developed under research limitations imposed by the Coronavirus pandemic, and so our previously existing data from [6] was used.

Successfully selected commands met the input duration requirements (short ≤ 0.5 s, long > 0.5 s) and the rest between inputs did not exceed a timeout (> 0.5 s). These values were selected from pilot studies and may not be optimal in general or for individuals. The classification accuracy calculated with these predetermined values was considered the Control. We then recalculated the classification accuracy for two additional conditions that included certain unsuccessful commands: 1) No Timeout and 2) Individual Classification. For the No Timeout condition, the first two inputs were classified according to the original input duration requirements, but without regard to the timeout (rest > 0.5 s, rest shown in Fig. 1A). The Individual Classification used the first two inputs as features for supervised machine learning and did not exclude

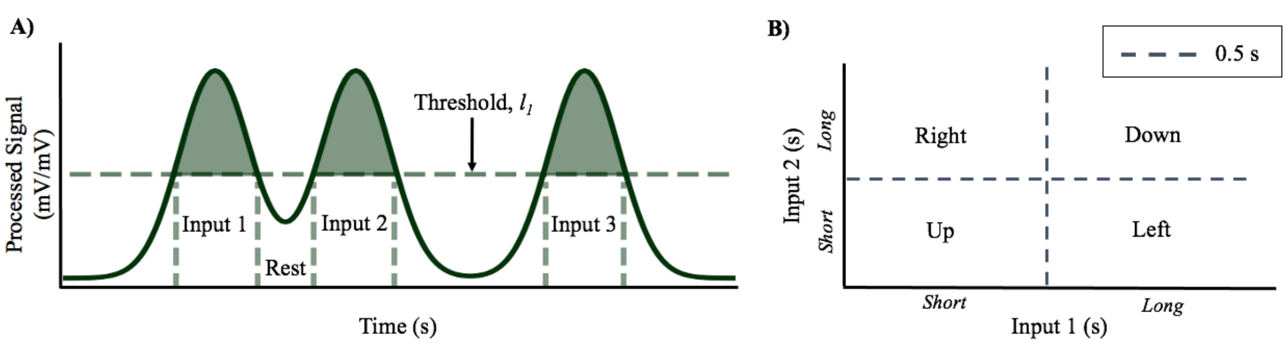

Figure 1. Overview of the coded sequence used in the sEMG method. A) An illustration of a processed EMG signal is shown here. The coded sequence requires two inputs above a threshold to select the command and the third input results in continuous, forward motion until the signal drops below the threshold. Subjects must rest (i.e., drop below the threshold) between inputs, but a rest greater than 0.5 s resets the sequence. B) The commands are defined by the combination of short (≤ 0.5 s) and long (> 0.5 s) inputs. For example, the "down" command is two long inputs.

any observations that exceeded the original timing requirements. (Refer to Table 1 for a summary of the Conditions with respect to the timing requirements.)

In the Individual Classification, we made several decisions to accommodate the limited number of observations per subject. The K Nearest Neighbors (KNN) algorithm was selected due to its straightforward implementation without the need to tune many parameters. The Python package *scikit-learn* [9] was used and the parameters for *k* and *weights* were set to 4 and "distance" (voting weighted by distance), respectively. The number of neighbors, *k*, was selected based on the number of observations per class and similar to the general rule ($\sqrt{observations} \approx$ 5-6). We assumed that subjects would perform command selection with precision, therefore we assigned more weight to closer neighbors. Typically, a 5- or 10-fold cross-validation strategy is used but based on the number of observations per class and the decision to use stratified folds, we implemented a 4-fold cross-validation with 3 iterations. Each fold was unique, and we used a train/test split of 80/20. The number of observations in the training set ranged from 29-32, but the testing set remained constant at 8 observations (2 per class). One subject had a total of 24 useable observations, which resulted in a 2-fold cross-validation with 6 iterations and 4 neighbors. Unusable observations occurred when the subjects' attempts contained one or no inputs and accounted for 1.51% of the total.

We compared the average classification accuracy across the Control, No Timeout, and Individual Classification conditions and hypothesized that removing the predetermined timing requirements would yield better accuracy. We also calculated the average increase in accuracy from the Control to No Timeout, and the No Timeout to Individual Classification. The reserved test dataset was used for another Individual Classification accuracy estimate. Results are reported as ($\mu \pm \sigma$) unless shown otherwise.

III. Results

We ran a two-factor mixed model with a between-subjects factor of Group and a within-subjects factor of Condition. Significant effects ($p < 0.05$) were analyzed with the Tukey Honest Significant Difference test with the Satterthwaite method to calculate the degrees of freedom. The main effect of Condition was significant ($F(2,88) = 31.40, p < 0.0001$), but not Group ($F(3,44) = 1.86, p = 0.15$) nor the interaction between Group and Condition ($F(6,88) = 0.83, p = 0.55$). The Tukey test for Condition showed significant differences for each pairwise comparison. Classification accuracy improved from the Control to No Timeout ($p < 0.0001$), Control to Individual Classification ($p < 0.0001$), and from the No Timeout to Individual Classification ($p = 0.04$). The Conditions achieved average classifications accuracies of 66.15 ± 26.31% for the Control, 76.56 ± 21.52% for No Timeout, and 81.58 ± 20.30% for Individual Classification.

TABLE I. Timing Requirements per Condition

Condition	Timeout	Input Duration
Control	Yes	Yes
No Timeout	No	Yes
Individual Classification	No	No

Classification accuracy results varied as some subjects scored 100% accuracy in the Control and did not benefit from the other Conditions, whereas other subjects had large performance gains with loosened timing requirements. For No Timeout, the increase in percentage points ranged from 0.00 to 50.00 points with an average increase of 10.42 ± 13.71 points. The Individual Classification inherently included the No Timeout consideration. The change in classification accuracy percentage points from No Timeout to Individual Classification ranged from -12.29 to 36.46 points with an average increase of 5.02 ± 10.29 points. For the cases where the accuracies decreased from No Timeout (N = 17), the Individual Classification accuracies were either worse (N = 8), better (N = 8), or the same (N = 1) as compared to the Control.

The reserved test dataset provided another classification accuracy estimate. The resulting accuracy across groups was 85.00 ± 18.36%. These results were similar to the accuracy estimated from the training dataset and cross-validation.

IV. Discussion

The results supported our hypothesis and indicated that the majority of subjects could potentially benefit from customized adaptation in our sEMG command system. The main effect of Group was found to be not significant; therefore the training history did not appear to be a factor in the outcome of the conditions, and was consistent with our prior results of no significant performance differences after training [6].

The No Timeout condition either benefited (N = 33) or did not affect (N = 15) the classification accuracy. Within the combined Test 2 and 3 dataset across groups, the timeouts accounted for 16.98% of the total attempts. It was interesting to note that by Group timeouts occurred in 15.63%, 5.42%, 17.92%, and 28.96% of the attempts for the Repetition, Concurrent Feedback, Terminal Feedback, and Adaptive Threshold groups, respectively. The Concurrent Feedback group may have been more attuned to the timing because the visual aid inherently gave information about the timing between inputs. The Adaptive Threshold group may have struggled learning the timing with the changing threshold. The timeouts had a mean of 0.41 s and median of 0.31 s. For an online implementation, the timeout could be modified to the subject's maximum timeout or 95th percentile from the Command Accuracy Test.

The Individual Classification improved the estimated classification accuracy for most subjects. Figs. 2 through 4 display the input durations performed for the target commands, which have a similar layout to Fig. 1B with lines indicating the predetermined input duration. The subject's data shown in Fig. 2 demonstrates the benefit of Individual Classification when command inputs are precise but stray from the prescribed timing. In contrast, some subjects' estimated classification accuracies remained low or decreased. One possible explanation may be that some lower performing subjects inconsistently produced inputs, such that the classes are not easily separable, and the nearest neighbors mostly belong to other classes (see Fig. 3). Alternatively, high performers may produce inputs within the prescribed timing, but occur close to other classes (see Fig. 4).

In our study, the original classification accuracy ranged from 5.00% to 100.00% with low and high performers within

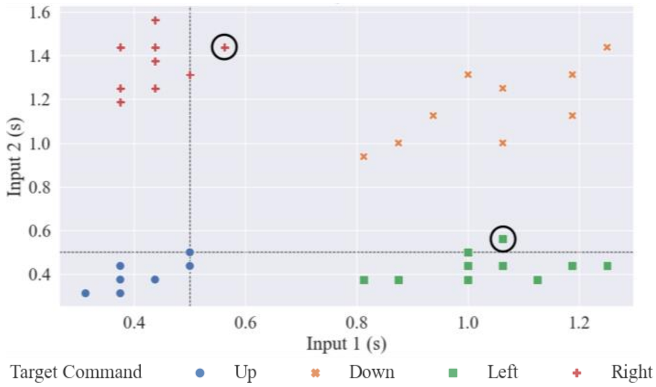

Figure 2. Example of a higher performing subject with classification accuracies of 95% (Control), 95% (No Timeout), and 100% (Individual Classification). Encircled points show attempts that did not have the correct input duration but remained close to their classes.

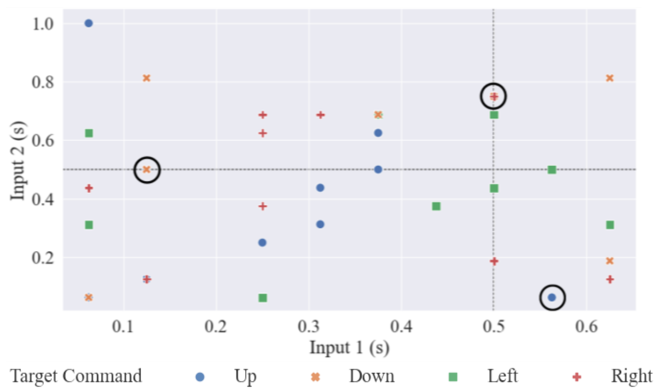

Figure 3. Example of a lower performing subject with classification accuracies of 30% (Control), 35% (No Timeout), and 41.67% (Individual Classification). Encircled points show attempts that did not have the correct input duration and/or were surrounded by other classes.

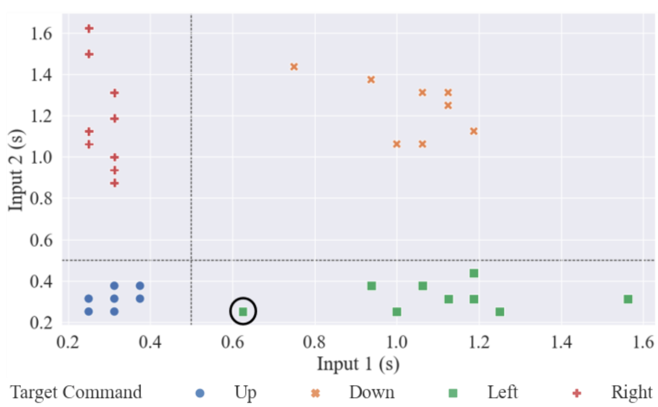

Figure 4. Example of a higher performing subject with classification accuracies of 100% (Control), 100% (No Timeout), and 96.88% (Individual Classification). Encircled point had the correct input durations but was closer in distance to the Up class causing a misclassification.

the subjects. Classifiers and adaptive features may improve the average performance but should target individuals that would benefit from customized timing requirements. For example, the subject whose data were depicted in Fig. 2 did not benefit from removing the timeout restriction but improved with Individual Classification. For Individual Classification, the average classification accuracy of 81.58% was comparable to Benchabane et al. [5] (87.8% average accuracy). Also, eleven of our subjects achieved classification accuracies of greater than 95%, similar to the LDA (average accuracy of 95.12%) and ALDA (95.45% to 100.00%) classifiers [2]. The results suggest that a customized approach to implementing classifiers and adaptive features could improve performance of already trained subjects. This customization could occur after a training period using the data from a Command Accuracy Test to determine the best approach for the subsequent online implementation.

This work used data already collected from subjects in an experiment not designed to test this hypothesis. The two main limitations of this study were the small dataset and post-data collection analysis. The maximum of 40 total observations was smaller than suggested by other researchers, who recommend starting with 75 observations per class [10]. The limited dataset for individuals also prevented a robust model selection step, but analysis done at the group level indicated a preference for KNN and selected parameters. Overall, the classification accuracy estimates indicated the potential benefits for individual subjects.

These are *potential* benefits. It is unclear whether an online adaptive sEMG command system that allows for custom timing would yield improved accuracy. How the subject would adapt in real-time is also unknown. Another possibility would be to use the adaptive component to train the subjects to become more efficient (e.g., smaller command time). Overall, these results are encouraging to pursue a future study to address these questions.

References

[1] I. Kyranou, S. Vijayakumar, and M. S. Erden, "Causes of Performance Degradation in Non-invasive Electromyographic Pattern Recognition in Upper Limb Prostheses," *Front Neurorobot*, vol. 12, p. 58, 2018, doi: 10.3389/fnbot.2018.00058.

[2] 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 *2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, 2013, pp. 4267–4270.

[3] 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, doi: 10.1109/TNSRE.2015.2401134.

[4] D. Yeung, D. Farina, and I. Vujaklija, "Directional Forgetting for Stable Co-Adaptation in Myoelectric Control," *Sensors*, vol. 19, no. 9, p. 2203, 2019.

[5] S. I. Benchabane, N. Saadia, and A. Ramdane-Cherif, "Novel algorithm for conventional myocontrol of upper limbs prosthetics," *Biomed. Signal Process. Control*, vol. 57, p. 101791, 2020.

[6] S. M. O'Meara, J. Karasinski, C. L. Miller, S. Joshi, and S. K. Robinson, "Effects of Augmented Feedback and Motor Learning Adaptation on Human-Automation Interaction Factors," *J. Aerosp. Inf. Syst.*, to be published.

[7] K. R. Lyons and S. S. Joshi, "Real-time evaluation of a myoelectric control method for high-level upper limb amputees based on homologous leg movements," 2016, pp. 6365–6368.

[8] S. M. O'Meara, M. C. Shyr, K. R. Lyons, and S. S. Joshi, "Comparing two different cursor control methods which use single-site surface electromyography," in *2019 9th international IEEE/EMBS conference on neural engineering (NER)*, 2019, pp. 1163–1166.

[9] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[10] C. Beleites, U. Neugebauer, T. Bocklitz, C. Krafft, and J. Popp, "Sample size planning for classification models," *Anal. Chim. Acta*, vol. 760, pp. 25–33, 2013.