

A Framework for Analyzing Continuous Transitions in Pattern Recognition-based Myoelectric Control

Shriram Tallam Puranam Raghu¹, Dawn T. MacIsaac¹, and Erik J. Scheme², *Senior Member, IEEE*

Abstract—Despite continued efforts to improve classification accuracy, it has been reported that offline accuracy is a poor indicator of the usability of pattern recognition-based myoelectric control. One potential source of this disparity is the existence of transitions between contraction classes that happen during regular use and are reported to be problematic for pattern recognition systems. Nevertheless, these transitions are often ignored or undefined during both the training and testing processes. In this work, we propose a framework for analyzing the transitions that occur during the voluntary changes between contraction classes during continuous control. As part of the framework, we propose new metrics that quantify the common types of errors that occur during transitions and compare them to existing metrics that apply only to the steady-state portions of the data. We then use this framework to analyze transition characteristics of 6 commonly used classifiers on a novel dataset that includes continuous transitions between all combinations of seven different contraction classes. Results show that a linear discriminant classifier consistently outperforms other conventional classifiers during both transitions and steady-state conditions, despite having an almost identical offline performance. Results also show that, although offline training metrics correlate with steady-state performance, they do not correlate with transition performance. These insights suggest that the proposed framework could provide a different and more representative picture of a classifier’s performance, potentially narrowing the gap between offline performance and online usability.

Index Terms—Surface Electromyography (SEMG); SEMG pattern recognition; evaluation metrics; transition metrics; usability; offline accuracy.

I. INTRODUCTION

Pattern recognition (PR) of surface electromyography (SEMG) signals is an active research area in several domains due to the wealth of information it provides. One of the most common applications of SEMG-PR is myoelectric control, such as is used to control prosthetic devices [1], [2]. Compared to conventional EMG control of prostheses, which typically uses only the amplitude information from one or two SEMG signals, PR-based systems leverage more information from multiple signals by extracting temporal and spatial features, enabling the discrimination of more movements [3]. SEMG-PR has also been explored for use in rehabilitation [4], and more recently, human-computer interaction (HCI) based applications such as gesture recognition [5] and computer game interaction [6].

*This work was supported by the Natural Sciences and Engineering Research Council of Canada (DG 2014-04920).

¹All authors are with the Department of Electrical and Computer Engineering, University of New Brunswick, Canada, stallam@unb.ca, dmac@unb.ca, escheme@unb.ca

²E. Scheme is also with the Institute of Biomedical Engineering, University of New Brunswick, Canada, escheme@unb.ca

Despite the advantages of SEMG-PR, several factors have been shown to significantly affect the performance of SEMG-PR systems [7], [8]. These include confounding influences such as limb position, electrode shift, fatigue, changes in patterns over time, and in general, noisy decision streams resulting from the inherently stochastic EMG source [9]. Many groups have endeavored to improve the robustness of SEMG-PR systems using a variety of approaches. Some have designed robust feature sets [10], [11], whereas others have employed filtering techniques to improve the quality of the underlying SEMG signal [12]. Post-processing techniques such as majority vote (MV) [3] or rejection [13] have also been explored to improve the quality of the decision stream. Recently, temporal deep learning algorithms such as Recurrent Neural Networks (RNNs) and Temporal Convolutional Networks (TCNs) have achieved high classification accuracies by exploiting the temporal information in the EMG signals [14], [15]. These algorithms have been shown to outperform traditional classifiers as well as more spatial deep learning algorithms such as convolutional neural networks (CNNs) and thus, are being actively researched for use in SEMG-PR systems.

The relatively high accuracies reported in these and other SEMG-PR studies are encouraging, but research has also shown that offline accuracy does not generally predict online usability [16], [17]. Consequently, many researchers now employ usability tests such as a Fitts law test [18], [19] to test the benefits of proposed techniques. While it is of great importance to measure performance during online usage, the value in these tests is limited to doing just that. They offer little insight into the apparent disconnect with offline accuracy or any possible mechanisms for observed improvements [20]. Because the user is ‘in-the-loop’ (responding to controller-specific feedback), they also preclude the subsequent use of collected data to compare or evaluate other control schemes.

Some studies have pointed to the ambiguous transition period between contractions, in which no pre-trained class may be represented, as a possible explanation for disagreement between offline accuracy and online usability [21], [22]. Indeed, most offline studies collect static and/or separate sets of motions as training and testing data, from which classifiers are built and evaluated. However, the practical use of SEMG-PR control systems requires continuous and dynamic transitions from one contraction to another. Training and evaluating classifiers using disjointed contractions may, therefore, not capture information that explains or accounts for the dynamics seen during usability.

Hudgins et al., in their seminal paper [11], acknowledged

the involvement of dynamics by training with ramp contractions, and others have since followed [17], [23]. Increased usability has been observed when training the classifier with ramp contractions as opposed to static contractions, despite a corresponding decrease in offline classifier accuracy. Scheme et al. explicitly evaluated the impact of contraction dynamics, and their results emphasized the importance of incorporating dynamics in the training process [22].

Although this has led to notable improvements, ramp contractions only model a subset of the dynamics - transitioning from rest to active contraction within a single motion class. Few studies have explicitly included or analyzed transitions from one motion class to another as part of these dynamics [24]. Incorporating all transitions during the training process may not be clinically practical given the large number of possible transition combinations (from each type of contraction to every other type), but an understanding of how these transitions impact current control schemes is warranted. Studying these transitions, however, is not straightforward.

When training and testing with static contractions to represent a motion class, segmentation of data is often trivialized by the collection process (stationary contractions are held for long periods and separated by a period in a neutral position). However, the segmentation, labeling, and evaluation of continuous data that include transitions pose a challenge, as it is not always easy to infer the start or endpoints of the motion classes [25]. While inertial measurement units [26], motion gloves, and motion capture systems [27] may be used to capture this intent, each approach has its own set of limitations and added complexity when trying to synchronize and align with corresponding EMG data [28]. Furthermore, differences in the relationship between the kinetics and kinematics observed by these systems during eccentric and concentric contractions further complicate the timing relationship with EMG [29].

In part because of these challenges, transition dynamics have typically been ignored during training and offline evaluation of SEMG-PR. As a result, transitions continue to be problematic during online control [30], and their influence on usability remains unclear. Since there is currently no established framework available to analyze continuous EMG transition data offline, in this work, we propose a framework to fill this gap. The framework includes six performance metrics designed specifically to quantify the common types of errors that occur during transitions. We then use this novel framework to compare the performance of six classifiers commonly used in SEMG-PR, during continuous transitions between 7 different motion classes.

II. METHODS

A. Data Acquisition

Data were collected from ten able-bodied participants recruited from a graduate student population. All participants gave informed consent as approved by the University of New Brunswick's Research Ethics Board (REB #2015-134). Eight bipolar channels of SEMG signals were recorded using the custom UNB smart-electrode cuff [31]. Data were sampled at

1 kHz with a 12-bit resolution, and electrode pairs were placed equidistantly around the participant's dominant forearm, proximal to the elbow. The signals were recorded on a desktop computer running a MATLAB® based data collection program [32]. Signals were digitally filtered to remove any power-line interference using a 3rd order Butterworth band-stop filter at 60 Hz, 180 Hz, and 300 Hz.

B. Training and Testing Protocol

Four sets of conventional training repetitions were recorded from each participant. Each set comprised a repetition of each contraction class, beginning from a neutral position and ramping up to a moderate level of contraction. Three-second repetitions of seven classes were included: No Movement (NM), Wrist Flexion (WF), Wrist Extension (WE), Wrist Pronation (WP), Wrist Supination (WS), Chuck Grip (CG) and Hand Open (HO). Participants were prompted by a static image on a computer monitor to initiate each contraction and were guided for pace by a 3-second countdown timer. An additional 3-seconds were given at the end of each contraction to give participants time to return to a neutral position, rest, and prepare for the next contraction. Data were collected only during the prompted contractions and used to train the classifiers outlined below.

Three sets of continuous transition test data were also collected. In these test sets, participants were prompted using the same visualizations as before, but with no delay between motions. This required them to directly transition from one class to another without intentionally returning to a neutral position between transitions, and without any knowledge of the next class. When the participants completed a transition from one class to the next, they were instructed to maintain that contraction (in steady-state) until the next class was prompted. At that time, they were asked to change to the new class immediately. Each set began in the no motion class and included all combinations of transitions from each class to every other class. Prompts were changed every 3 seconds and, other than ensuring that all combinations were included, transitions were randomly ordered across sets and participants so that participants could not anticipate the next class. No feedback was provided to the participants, and they were encouraged to perform the contractions as naturally as possible. The SEMG signals were recorded continuously throughout each full set. A section of SEMG signals obtained from a set of representative contractions is shown in Figure 1, with the prompted contraction and timing shaded in different colours.

All signals were segmented with a frame length of 160 ms and an increment of 16 ms for classification. Hudgins time-domain feature set [11] (Mean Absolute Value (MAV), Slope Sign Changes (SSC), Waveform Length (WL) and Zero Crossings (ZC)) was computed for each frame per channel, resulting in a total of 32 features (4 features * 8 channels) per frame.

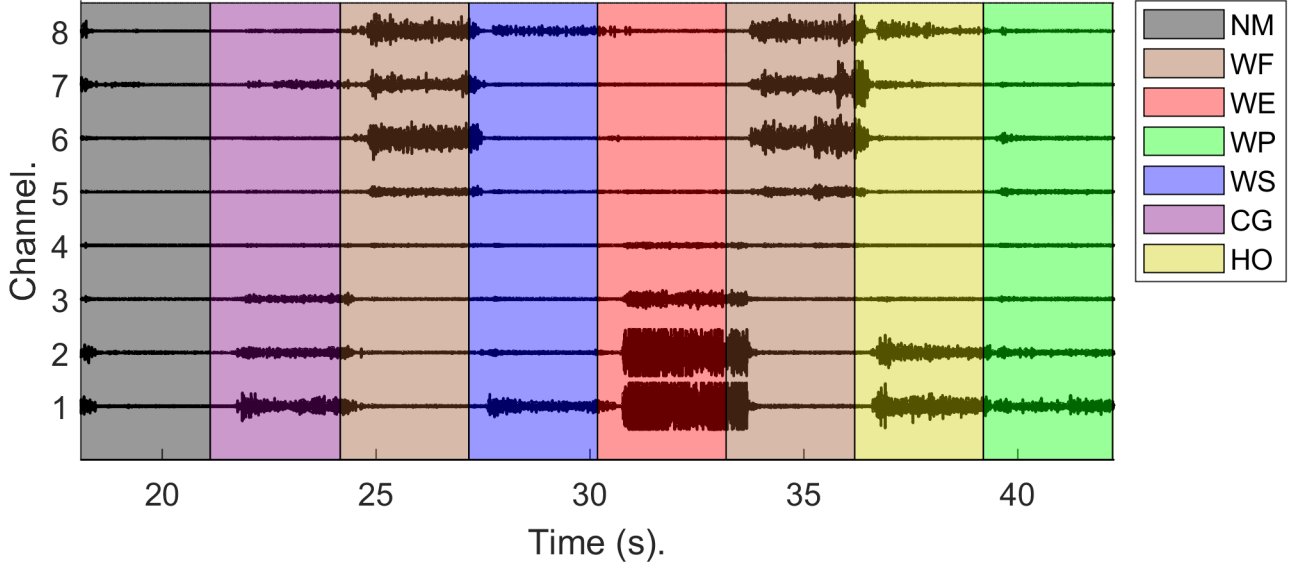


Fig. 1. A 24 second SEMG time series segment obtained from a test set. The shaded regions indicate the prompted contraction.

C. Comparing Classifier Performance

The performance of several commonly used classification techniques was compared in this work. These included Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Support Vector Machine with Linear Kernel (SVM), K-Nearest Neighbours (KNN), Multilayer Perceptron-Artificial Neural Network (MLP-ANN) and Random Forest (RF) [1], [33]. LDA was implemented as described in [13], SVM was implemented using LIBSVM (version 3.23), MLP-ANN was implemented using MATLAB® Deep Learning Toolbox, and the rest of the classifiers were implemented using the MATLAB® Statistics and Machine Learning Toolbox. The MATLAB® default settings were used in the implementation of the MLP-ANN and RF. MATLAB® 2019a was used to conduct the analysis. The hyperparameters of SVM, KNN, MLP-ANN, and RF were tuned using data from two randomly chosen participants using 10-fold cross-validation. Based on these pilot results, a single value was chosen for each hyperparameter and used for the remainder of the experiment. The chosen hyperparameters are listed in Table I.

Classifier	Hyperparameters
SVM	$c = 1$
KNN	$k = 5$
MLP-ANN	1 Hidden Layer with 5 Neurons
RF	100 Learning Cycles

TABLE I
FEATURE SETS USED IN THIS STUDY.

An essential aspect of supervised machine learning is the creation of labels to train and evaluate classifier performance. In continuous transition data, however, this process is not trivial. Labels can readily be assigned to the steady-state portion of signals, but the labeling of frames within the

dynamic portions of contraction may depend on the previous class, the reaction time of the user, the speed of transition, and the kinematics of the contraction. Additionally, because multi-degree of freedom joint control is continuous, the use of discrete classes may lead to undefined periods during these transitions [34]. For these reasons, in this work, the transition and steady-state regions were analyzed separately.

To identify the steady-state regions, which most classifiers have been shown to accurately classify [17], each classifier's decision stream was smoothed by taking a 9-frame majority vote across the stream. The beginning of steady-state was then defined as the first frame after a prompt whose classification decision corresponded to the prompted class. In the rare instance when no corresponding frames were identified during a prompt, it was assumed that an operator error had occurred (e.g. the participant elicited the wrong contraction), and the contraction was discarded from the analysis. The same process was followed to find the end of steady state, which always occurred shortly after the prompt for the upcoming contraction. The end frame for each steady-state period was defined as the first frame after a prompt whose classification decision did not correspond to the previous prompt class. Finally, to compensate for the delay introduced by the majority vote, all steady-state labels were adjusted backward by 4 frames.

The start and end frames of the steady-states were then used to label the frames. All frames that fell between a given start and endpoint were assigned the label corresponding to the prompt. All frames that fell between consecutive steady-states were then defined as transitions and were not assigned labels. Because these frames had no labels, traditional accuracy metrics could not be applied to transition regions, and so a new set of performance metrics were developed. This new framework for assessing classifier performance during both

steady-state and in transition is described in the next section.

D. Assessing the performance of the classifiers

Leave-one-out cross-validation accuracy, commonly used to assess offline classifier performance in SEMG-PR studies [35], was used to perform a participant-wise leave-one-set-out analysis of the training data. This was done to establish a baseline that is comparable with previous literature and to confirm that the classifiers were properly trained. In each fold, a classifier was trained using all but one training set and tested with the remaining set. This process was repeated until all the sets were assessed, and the average error across all sets and participants was recorded as the offline training error. This process was completed for each of the six classifiers and the results were analyzed for significant differences with a Kruskal-Wallis test as the distributions of offline training errors failed normality testing.

Next, the test sets were used to evaluate each classifier's steady-state and transition performance. The developed framework facilitates the comparison of classifier performance across the same set of steady-state contractions and transitions. Because each classifier is evaluated using the same data, unlike conventional usability tests (where each system must be tested separately due to the influence of user-in-the-loop control), results can be compared between classifiers, alleviating the need for a deterministic ground-truth. That is to say, there is value in knowing which classifiers respond more quickly to changes and with less volatility than others.

1) *Steady-state metrics*: Commonly used metrics were used to assess the steady-state regions as described below [36].

- a) **Total Error Rate (TER)**: the percentage of incorrect decisions in a steady-state region. Lower values are better.
- b) **Active Error Rate (AER)**: The percentage of decisions corresponding to an active (i.e. any class besides no motion) but incorrect class. Lower values are better.
- c) **Instability (INS)**: the number of times, in a steady-state region, that a pair of consecutive active class decisions differs, normalized by the total number of decisions in that region. Lower values are better.

2) *Transition metrics*: In total, six transition metrics were employed to measure the transition performance as described below. The first three metrics focus on the response times of the classifiers. The last three focus on fluctuations within transition regions. A common phenomenon during transitions that degrades usability is the rapid and stochastic fluctuation between classifier decisions [20]. As in the steady-state metrics, instability captures these rapid changes, and two related metrics, tertiary class error, and percent no movement, help to further discern how these changes manifest.

- a) **Offset delay (T_{OFFSET})**: the delay between a change in prompt and when the classifier leaves the previous steady state. Lower values are better.
- b) **Onset delay (T_{ONSET})**: the delay between a change in prompt and when a classifier reaches the corresponding steady-state. Lower values are better.

- c) **Transition Duration ($T_{TRANSITION}$)**: the duration between two consecutive steady states. Lower values are better.
- d) **Instability (INS)**: the number of times, in a region of transition, that a pair of consecutive active class decisions differ, normalized by the total number of decisions in that region. Lower values are better.
- e) **Tertiary Class Error (TCE)**: the number of decisions in a transition region that do not correspond to the steady states adjacent to the region, or to NM, normalized by the total number of decisions in that region. TCE can be considered similar to AER, but for transitions. Lower values are better.
- f) **Percent No Movement (PNM)**: the number of NM decisions in a transition region, normalized by the total number of decisions in that region. Higher values are better.

Figure 2 depicts the transition metrics for an example decision stream. The figure shows a typical transition from one active class to another. In the figure, each data point represents a colour coded confidence in the classification decision for a frame. The colour of the data point reflects the class decision. The lower shaded regions indicate the prompted classes by colour (with a change in prompt at frame number 6973). The upper shaded regions indicate the assigned labels by colour (obtained as explained in Section II-C). When the colour of the data points matches the colour of the upper shaded regions, the classification decision matched the frame label. Data in the example figure were produced using an LDA classifier, but the figure framework, in general, is representative of all classifiers.

These metrics were computed and averaged across the three continuous test trial sets on a per participant and per transition basis. The transitions were then grouped into three categories: transitions from an active class to rest (A2R), transitions from rest to an active class (R2A), and transitions from one active class to another (A2A). This process was completed for each of the six classifiers and the results were analyzed for significant differences between groups and classifiers with using a Kruskal-Wallis test. When significant differences were indicated, a post-hoc pairwise multiple comparison test with Dunn-Šidák correction was applied. Alpha values were always set to 0.05. Finally, a Pearson's correlation analysis was performed to ascertain if the offline TER correlated to any of the test framework metrics.

III. RESULTS

The training errors are provided for the different classifiers in Tables II. Results show that all classifiers performed comparably in terms of TER on the training set. RF obtained the lowest TER of 12.7%, but no significant differences were found ($p > 0.9$). The errors obtained were also within the reported ranges in the wider literature [22], suggesting that the classifiers were trained appropriately.

Table III shows the steady-state metrics obtained from the continuous test set. Unlike in training, these results show that LDA was the best classifier across all three metrics, while

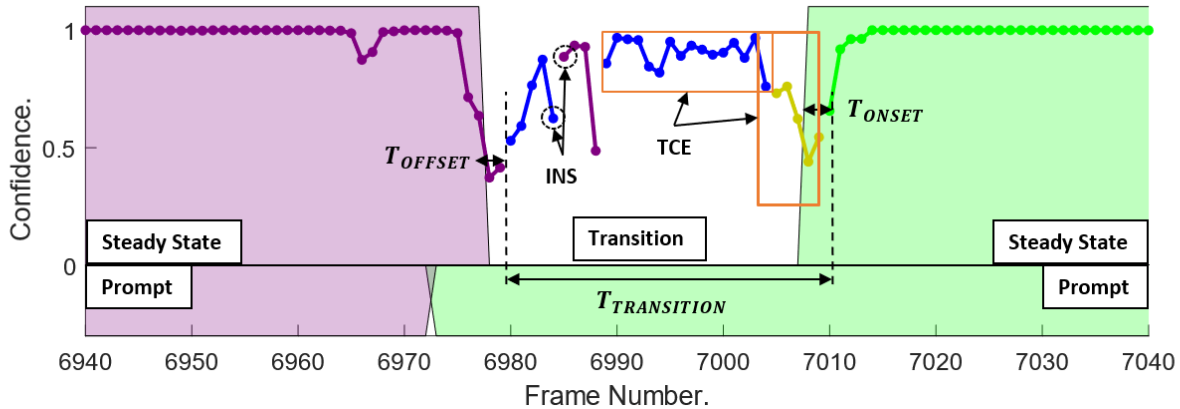


Fig. 2. Visual depiction of transition metrics for an example decision stream.

Classifier	Training TER (%)
ANN	15.1 ± 8.6
KNN	15.8 ± 8
LDA	14.1 ± 7.3
QDA	14.1 ± 7.8
RF	12.7 ± 6.5
SVM	14.5 ± 8.1

TABLE II
COMPARISON OF OFFLINE TRAINING ERROR (%).

Classifier	AER (%)	TER (%)	INS (%)
ANN	14.7 ± 10.3	19.3 ± 13.4	3.0 ± 2.2
KNN	16.2 ± 11.1	20.1 ± 13.2	3.9 ± 2.9
LDA	11.2 ± 9.3⁺	18.7 ± 13.8	2.0 ± 1.9⁺⁺
QDA	18.1 ± 14.2	21.3 ± 15.7	3.3 ± 2.9
RF	15.4 ± 11.9	20.4 ± 14.7	3.5 ± 2.7
SVM	14.2 ± 10.5	19.2 ± 14.0	3.0 ± 2.4

TABLE III
COMPARISON OF STEADY STATE METRICS FOR DIFFERENT CLASSIFIERS (MEAN AND STANDARD DEVIATION). BOLDED VALUES INDICATE BEST PERFORMERS. ++INDICATES SIGNIFICANTLY BETTER THAN ALL OTHER CLASSIFIERS; +INDICATES SIGNIFICANTLY BETTER THAN KNN AND QDA.

QDA had the worst AER and TER, and KNN had the worst INS. Despite the offline training error being almost identical for LDA and QDA, the differences in the steady-state metrics were substantive. Statistical differences were found across classifiers for AER and INS ($p < 0.05$) but not TER ($p > 0.9$). The post-hoc analysis indicated that LDA had significantly lower INS than all other classifiers and lower AER than QDA and KNN ($p < 0.05$ in both cases). No other significant differences were noted.

When analyzing the transition metrics, first, a preliminary analysis was done to compare the metrics from the three transition groups, pooling data from all the classifiers and participants. The results indicated that $T_{TRANSITION}$, INS, and TCE were smaller for the R2A case compared to the other cases. The A2R case was second-best, with slightly lower $T_{TRANSITION}$ and INS, and higher PNM, compared to the A2A group. A Kruskal-Wallis test yielded statistically

significant differences across the transition groups ($p < 0.001$) and the post-hoc pairwise multiple comparison tests confirmed the observation that R2A outperformed both other groups and A2R outperformed A2A ($p < 0.05$). These results indicate that transitions from rest to active classes are handled best by the classifiers, followed by transitions from active to rest. These results were also found to be consistent across each type of classifier.

Next, for each of the transition groups, a comparison of classifier performance was conducted. The results of each of these comparisons are shown in Tables IV, V and VI, for R2A, A2R, and A2A respectively. The results obtained show that no single classifier performed best across all metrics in all 3 groups, but that LDA stood out in many regards. First, LDA yielded the best performance for all metrics in the A2R group. A Kruskal-Wallis test indicated that there were statistical differences between the classifiers for each metric ($p < 0.05$) except for T_{OFFSET} ($p > 0.1$). For those metrics yielding significant differences, post-hoc pairwise multiple comparison tests indicated that LDA significantly outperformed all other classifiers ($p < 0.05$).

In the R2A group, a Kruskal-Wallis test yielded statistical differences across the classifiers ($p < 0.05$) for all metrics except for T_{ONSET} ($p > 0.25$) and PNM ($p > 0.8$). For $T_{TRANSITION}$, INS, and TCE, LDA again yielded the best results and post-hoc pairwise multiple comparison tests confirmed that the LDA outperformed QDA in $T_{TRANSITION}$, KNN in INS, and QDA and KNN in TCE ($p < 0.05$). However, QDA was significantly better than LDA in T_{OFFSET} ($p < 0.05$). No other classifier outperformed any other on more than one metric.

Finally, in the A2A group, LDA was again the best performer in terms of T_{OFFSET} , INS, TCE, and PNM. A Kruskal-Wallis test revealed significant differences for all of these metrics ($p < 0.05$) and the post hoc analysis revealed that LDA outperformed all other classifiers ($p < 0.05$). Significant differences were also found in $T_{TRANSITION}$ and T_{ONSET} for this transition group, and unlike the previous results,

Classifier	T_{OFFSET} (ms)	T_{ONSET} (ms)	$T_{TRANSITION}$ (ms)	INS (%)	TCE (%)	PNM (%)
ANN	444.2 ± 124.3	504.2 ± 138.0	60.0 ± 41.6	2.9 ± 4.8	10.0 ± 13.6	0.3 ± 1.4
KNN	441.7 ± 127.4	527.5 ± 169.8	82.2 ± 74.6	6.6 ± 8.5	17.8 ± 20.0	0.4 ± 2.3
LDA	507.6 ± 138.8	561.0 ± 144.6	53.4 ± 36.2⁺	2.0 ± 3.7⁺	7.3 ± 12.1⁺	0.3 ± 2.1
QDA	444.1 ± 156.7	504.7 ± 169.7	108.9 ± 108.8	5.0 ± 6.6	21.4 ± 24.1	0.3 ± 2.2
RF	450.8 ± 135.6	534.7 ± 204.5	84.0 ± 116.0	3.3 ± 6.8	12.3 ± 19.3	1.0 ± 5.0
SVM	451.3 ± 125.9	518.1 ± 135.2	66.8 ± 52.9	3.9 ± 6.9	11.3 ± 16.8	0.6 ± 2.8

TABLE IV

COMPARISON OF TRANSITION METRICS ACROSS THE DIFFERENT CLASSIFIERS IN THE R2A TRANSITION GROUP (MEAN AND STANDARD DEVIATION). BOLDDED VALUES INDICATE BEST PERFORMERS. + INDICATES SIGNIFICANTLY BETTER THAN AT LEAST ONE CLASSIFIER, - INDICATES WORSE THAN AT LEAST ONE CLASSIFIER.

Classifier	T_{OFFSET} (ms)	T_{ONSET} (ms)	$T_{TRANSITION}$ (ms)	INS (%)	TCE (%)	PNM (%)
ANN	221.6 ± 195.6	456.3 ± 293.6	250.1 ± 237.3	4.3 ± 5.3	42.3 ± 31.8	55.7 ± 31.5
KNN	225.4 ± 216.5	453.2 ± 252.1	245.6 ± 229.3	6.1 ± 6.8	42.0 ± 29.0	55.2 ± 29.4
LDA	185.9 ± 173.8	230.4 ± 236.5⁺⁺	87.1 ± 96.4⁺⁺	1.5 ± 3.8⁺⁺	15.5 ± 21.8⁺⁺	83.8 ± 22.5⁺⁺
QDA	279.8 ± 204.9	649.3 ± 283.2	409.4 ± 307.4	6.0 ± 5.2	56.4 ± 29.0	39.2 ± 26.5
RF	202.4 ± 221.5	438.3 ± 243.1	252.9 ± 254.0	5.5 ± 6.4	41.7 ± 29.3	56.5 ± 29.9
SVM	210.5 ± 199.6	437.3 ± 274.3	238.4 ± 205.3	4.1 ± 4.7	42.0 ± 29.0	55.3 ± 29.5

TABLE V

COMPARISON OF TRANSITION METRICS ACROSS THE DIFFERENT CLASSIFIERS IN THE A2R TRANSITION GROUP (MEAN AND STANDARD DEVIATION). BOLDDED VALUES INDICATE BEST PERFORMERS. ++INDICATES SIGNIFICANTLY BETTER THAN ALL OTHER CLASSIFIERS.

Classifier	T_{OFFSET} (ms)	T_{ONSET} (ms)	$T_{TRANSITION}$ (ms)	INS (%)	TCE (%)	PNM (%)
ANN	163.2 ± 318.5	491.8 ± 217.2	331.9 ± 318.1	7.9 ± 6.5	32.6 ± 25.3	18.9 ± 24.0
KNN	175.0 ± 284.3	479.9 ± 221.5	325.5 ± 310.5	11.1 ± 8.2	35.9 ± 26.6	15.0 ± 20.1
LDA	150.6 ± 284.1	533.7 ± 198.4 ⁻⁻	404.5 ± 296.6 ⁻⁻	5.0 ± 5.4⁺⁺	19.9 ± 21.2⁺⁺	42.1 ± 26.4⁺⁺
QDA	221.0 ± 305.4	523.7 ± 297.5*	337.2 ± 334.6	9.3 ± 7.5	39.1 ± 28.2	7.6 ± 15.5 ⁻⁻
RF	184.0 ± 306.0	492.1 ± 240.4	336.1 ± 330.6	9.6 ± 8.7	33.3 ± 26.2	18.0 ± 22.8
SVM	174.0 ± 322.5	495.6 ± 226.8	339.7 ± 322.2	7.4 ± 6.4	32.2 ± 25.9	19.7 ± 24.8

TABLE VI

COMPARISON OF TRANSITION METRICS ACROSS THE DIFFERENT CLASSIFIERS IN THE A2A TRANSITION GROUP (MEAN AND STANDARD DEVIATION). BOLDDED VALUES INDICATE BEST PERFORMERS. ++INDICATES SIGNIFICANTLY BETTER THAN ALL OTHER CLASSIFIERS, - INDICATES SIGNIFICANTLY WORSE THAN ALL OTHER CLASSIFIERS, AND * INDICATES EXCLUDED FROM THE COMPARISON.

the post hoc analysis revealed that LDA performed worse compared to all other classifiers for $T_{TRANSITION}$, and all but QDA for T_{ONSET} ($p < 0.05$).

It should be noted that using the timing of the prompts as the ground truth for computing delay estimates introduces unrealistically long biases (they include the user response delay). While the relative performance of classifiers remains unaffected and valid, the absolute numbers should be interpreted carefully with this understanding.

Finally, results from the correlation analysis for the steady-state and transition metrics are shown in Tables VII and VIII respectively. The offline TER showed a moderate, positive, and statistically significant correlation with all of the steady-state metrics ($p < 0.001$). Conversely, no consistent correlation was found across the transition metrics. Offline TER showed a moderate, positive, and statistically significant correlation with TCE and PNM in the R2A group ($p < 0.001$), but no other metrics showed any significant correlation for that group ($p > 0.1$). Offline TER showed no significant correlation with any of the transition metrics in the A2R group ($p > 0.35$) and, for the R2A group, only INS showed a statistically significant (albeit weak) correlation ($p < 0.05$ for INS, $p > 0.1$ for all the other metrics).

Metric	R
AER	0.58*
TER	0.56*
INS	0.54*

TABLE VII

CORRELATION COEFFICIENT (R) BETWEEN OFFLINE TER AND STEADY-STATE METRICS. * INDICATES STATISTICALLY SIGNIFICANT ($p < 0.05$) VALUES.

Metric	R_{R2A}	R_{A2R}	R_{A2A}
T_{OFFSET}	0.15	0.03	0.12
T_{ONSET}	0.07	-0.12	0.19
$T_{TRANSITION}$	0.09	-0.02	-0.05
INS	0.21	0.09	0.26*
TCE	0.55*	0.1	0.22
PNM	0.46*	0.1	0.20

TABLE VIII

CORRELATION COEFFICIENT (R) BETWEEN OFFLINE TER AND TRANSITION METRICS ACROSS THE THREE GROUPS. * INDICATES STATISTICALLY SIGNIFICANT ($p < 0.05$) VALUES.

IV. DISCUSSION

In this work, we highlighted the need to analyze transitions when evaluating classifiers and proposed a novel framework and metrics. This simple framework can be used to quantify performance during transitions and steady-state regions of con-

tractions separately to provide a more complete picture of the benefits of a particular classifier. We then used the framework to compare 6 commonly used classifiers to determine how their performances differ between steady-state and transitions, and if a single classifier performs well across all metrics. We also compared these results to the commonly used leave-one-set-out offline training error of the classifiers.

The offline training error showed no significant differences across the classifiers, as has been shown in other works [37], [38]. These results, however, did not translate to either of the steady-state or transition metrics obtained from the continuous transition test data. For instance, even though LDA and QDA had almost identical offline training performance, when performance was evaluated on the test data, LDA was frequently significantly better across most metrics (in both transition and steady-state), while QDA was often the worst. These discrepancies add to the argument that offline performance is an incomplete indicator of the overall performance of the PR-system.

When evaluating transition errors, there were significant differences across the three different groups of transitions (R2A, A2R, and A2A). Classifiers consistently performed better on the R2A group compared to the other two groups when considering $T_{TRANSITION}$, INS and TCE. This is perhaps reflective of the inclusion of ramping contractions during training, which informs the classifiers of the dynamics of transition from rest to steady-state [22]. This group did yield a lower PNM compared to the other groups, which can sometimes indicate higher volatility, but only if this volatility is also supported by high INS and TCE. Since these metrics were also low for the R2A case, the low PNM may simply be a reflection of a low volatility transition.

The A2R transitions outperformed the A2A transitions in $T_{TRANSITION}$, INS, and TCE. Although the training data included only ramps from rest to active classes, this could still act to inform the classifiers about the boundaries between each class and rest. Conversely, the transitions between active contraction classes are not typically included during training because of the onerous number of necessary combinations (particularly for classifiers with many classes). These transitions may travel through the NM class (the first class is released before initiating the second) but may also involve traversing feature space through other class boundaries in undefined ways.

LDA is commonly used in the literature because it is simple to implement, fast to train, and reduces over-tuning [39]. Despite being touted for its simplicity, LDA was found to be the most promising classifier across all transition cases in this work. It had the lowest transition durations for the R2A and A2R groups, and, despite having the highest transition duration for the A2A group, LDA had significantly lower INS and TCE, and significantly higher PNM, indicating a less volatile transition compared to the other classifiers. Likely, the generative and parametric nature of the LDA helps it generalize better than other classifiers when the test scenario includes more variability than seen during training. Although

QDA is a similar generative classifier, it performs relatively poorly, even compared to the more discriminative classifiers. This may be a result of over-tuning because, unlike the LDA which pools covariance estimates, the QDA models each class separately. This supports that caution should be taken with more complex classifiers that may not generalize beyond training.

The pooling assumption used by LDA, however, is not always advantageous as it may cause issues when classes have vastly different covariances. For instance, the NM class tends to have a much smaller covariance as compared to classes that involve active contractions. Using the pooled covariance effectively increases the time it takes the LDA to transition out of NM. This is evidenced by the relatively large T_{OFFSET} in the R2A group. Fortunately, this effect is reversed when transitioning into NM, as supported by the small T_{ONSET} in the A2R case. The overall strong performance of LDA suggests that accepting this trade-off may lead to better generalization.

Further investigation into the transitions between different active classes indicated that all transitions are not necessarily equally classifiable. For instance, we found that transitions from any active class to wrist pronation resulted in double the TCE than those transitioning to wrist supination. We also noted that transitions to chuck grip were 5 times faster than those to wrist extension. These vast differences in performance and behaviours during transitions warrant the continued exploration of temporal and state-based approaches for myoelectric control.

It is important to note that, in this study, the effect of different features was not explored. This is an important consideration, as features have been found to have a critical impact on overall classification performance [37]. Here, we used the commonly cited Hudgins' TD feature set to focus on classifiers, but many features have now been shown to perform better in various circumstances [40]. Consequently, future work will explore how the suite of EMG features, either engineered or learning via deep learning, impact transition behaviour.

Finally, although the proposed metrics are capable of quantifying performance during transitions, it is not yet known if they will correlate with usability. As has been reported with online usability [41], in this work, we found no consistent correlation between offline accuracy and transition metrics across the transition groups. Specifically, A2R and A2A group showed weak to no correlation across all metrics, while two metrics in the R2A group (TCE and PNM) showed moderate correlation. While the lack of consistent correlation does not conclusively link transition performance and usability, it offers sufficient potential to warrant further investigation.

ACKNOWLEDGMENT

We would like to thank NSERC for the funding. Additionally, we would like to thank Kadie Wright for collecting the data used in this study.

REFERENCES

- [1] E. Scheme and K. Englehart, "Electromyogram pattern recognition for control of powered upper-limb prostheses: State of the art and challenges for clinical use," *JRRD*, vol. 48, no. 6, p. 643, 2011, doi: 10.1682/JRRD.2010.09.0177.
- [2] S. Amsuss, P. M. Goebel, Ning Jiang, B. Graimann, L. Paredes, and D. Farina, "Self-Correcting Pattern Recognition System of Surface EMG Signals for Upper Limb Prosthesis Control," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 4, pp. 1167–1176, Apr. 2014, doi: 10.1109/TBME.2013.2296274.
- [3] K. Englehart and B. Hudgins, "A robust, real-time control scheme for multifunction myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 7, pp. 848–854, Jul. 2003, doi: 10.1109/TBME.2003.813539.
- [4] Z. O. Khokhar, Z. G. Xiao, and C. Menon, "Surface EMG pattern recognition for real-time control of a wrist exoskeleton," *BioMed Eng OnLine*, vol. 9, no. 1, p. 41, 2010, doi: 10.1186/1475-925X-9-41.
- [5] Z. Chen, N. Zhang, Z. Wang, Z. Zhou, and D. Hu, "Hand Gestures Recognition from Multi-channel Forearm EMG Signals," in *Cognitive Systems and Signal Processing*, vol. 710, F. Sun, H. Liu, and D. Hu, Eds. Singapore: Springer Singapore, 2017, pp. 119–125.
- [6] L. E. Nacke, M. Kalyn, C. Lough, and R. L. Mandryk, "Biofeedback game design: using direct and indirect physiological control to enhance game interaction," in *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*, Vancouver, BC, Canada, 2011, p. 103, doi: 10.1145/1978942.1978958.
- [7] E. Scheme and K. Englehart, "On the robustness of EMG features for pattern recognition based myoelectric control; A multi-dataset comparison," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Chicago, IL, Aug. 2014, pp. 650–653, doi: 10.1109/EMBC.2014.6943675.
- [8] M. M.-C. Vidovic, H.-J. Hwang, S. Amsuss, J. M. Hahne, D. Farina, and K.-R. Muller, "Improving the Robustness of Myoelectric Pattern Recognition for Upper Limb Prostheses by Covariate Shift Adaptation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 9, pp. 961–970, Sep. 2016, doi: 10.1109/TNSRE.2015.2492619.
- [9] 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, doi: 10.3390/s20061613.
- [10] D. Tkach, H. Huang, and T. A. Kuiken, "Study of stability of time-domain features for electromyographic pattern recognition," *J NeuroEngineering Rehabil*, vol. 7, no. 1, p. 21, 2010, doi: 10.1186/1743-0003-7-21.
- [11] 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, doi: 10.1109/10.204774.
- [12] D. Gradolewski, P. M. Tojza, J. Jaworski, D. Ambroziak, G. Redlarski, and M. Krawczuk, "Arm EMG Wavelet-Based Denoising System," in *Mechatronics - Ideas for Industrial Application*, vol. 317, J. Awrejcewicz, R. Szewczyk, M. Trojnecki, and M. Kaliczyńska, Eds. Cham: Springer International Publishing, 2015, pp. 289–296.
- [13] E. J. Scheme, B. S. Hudgins, and K. B. Englehart, "Confidence-Based Rejection for Improved Pattern Recognition Myoelectric Control," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 6, pp. 1563–1570, Jun. 2013, doi: 10.1109/TBME.2013.2238939.
- [14] Y. Hu, Y. Wong, W. Wei, Y. Du, M. Kankanhalli, and W. Geng, "A novel attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition," *PLoS ONE*, vol. 13, no. 10, p. e0206049, Oct. 2018, doi: 10.1371/journal.pone.0206049.
- [15] J. L. Betthausen, J. T. Krall, R. R. Kaliki, M. S. Fifer, and N. V. Thakor, "Stable Electromyographic Sequence Prediction During Movement Transitions using Temporal Convolutional Networks," in *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, San Francisco, CA, USA, Mar. 2019, pp. 1046–1049, doi: 10.1109/NER.2019.8717169.
- [16] I. Kyranou, S. Vijayakumar, and M. S. Erden, "Causes of Performance Degradation in Non-invasive Electromyographic Pattern Recognition in Upper Limb Prostheses," *Front. Neurobot.*, vol. 12, p. 58, Sep. 2018, doi: 10.3389/fnbot.2018.00058.
- [17] 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 *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug. 2007, pp. 4842–4845, doi: 10.1109/IEMBS.2007.4353424.
- [18] L. Hargrove, L. Miller, K. Turner, and T. Kuiken, "Control within a virtual environment is correlated to functional outcomes when using a physical prosthesis," *J NeuroEngineering Rehabil*, vol. 15, no. S1, p. 60, Sep. 2018, doi: 10.1186/s12984-018-0402-y.
- [19] 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, doi: 10.1109/TNSRE.2012.2226189.
- [20] J. W. Robertson, K. B. Englehart, and E. J. Scheme, "Effects of Confidence-Based Rejection on Usability and Error in Pattern Recognition-Based Myoelectric Control," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 5, pp. 2002–2008, Sep. 2019, doi: 10.1109/JBHI.2018.2878907.
- [21] A. M. Simon, L. J. Hargrove, B. A. Lock, and T. A. Kuiken, "A Decision-Based Velocity Ramp for Minimizing the Effect of Misclassifications During Real-Time Pattern Recognition Control," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 8, pp. 2360–2368, Aug. 2011, doi: 10.1109/TBME.2011.2155063.
- [22] E. Scheme and K. Englehart, "Training Strategies for Mitigating the Effect of Proportional Control on Classification in Pattern Recognition-Based Myoelectric Control," *JPO Journal of Prosthetics and Orthotics*, vol. 25, no. 2, pp. 76–83, Apr. 2013, doi: 10.1097/JPO.0b013e318289950b.
- [23] K. Englehart, B. Hudgins, P. A. Parker, and M. Stevenson, "Classification of the myoelectric signal using time-frequency based representations," *Medical Engineering & Physics*, vol. 21, no. 6–7, pp. 431–438, Jul. 1999, doi: 10.1016/S1350-4533(99)00066-1.
- [24] U. Cote-Allard et al., "Deep Learning for Electromyographic Hand Gesture Signal Classification Using Transfer Learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 4, pp. 760–771, Apr. 2019, doi: 10.1109/TNSRE.2019.2896269.
- [25] J. W. Robertson, K. B. Englehart, and E. J. Scheme, "Effects of Confidence-Based Rejection on Usability and Error in Pattern Recognition-Based Myoelectric Control," presented at the *Myoelectric Controls/Powered Prosthetics Symposium*, Fredericton, Canada, 2017.
- [26] M. T. Wolf et al., "Decoding static and dynamic arm and hand gestures from the JPL BioSleeve," in *2013 IEEE Aerospace Conference*, Big Sky, MT, Mar. 2013, pp. 1–9, doi: 10.1109/AERO.2013.6497171.
- [27] G. N. Pradhan, N. Engineer, M. Nadin, and B. Prabhakaran, "Integration of Motion Capture and EMG data for Classifying the Human Motions," in *Proceedings of the 2007 IEEE 23rd International Conference on Data Engineering Workshop*, USA, Apr. 2007, pp. 56–63, doi: 10.1109/ICDEW.2007.4400973.
- [28] A. Ameri, E. J. Scheme, E. N. Kamavuako, K. B. Englehart, and P. A. Parker, "Real-Time, Simultaneous Myoelectric Control Using Force and Position-Based Training Paradigms," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 2, pp. 279–287, Feb. 2014, doi: 10.1109/TBME.2013.2281595.
- [29] D. Ewins and T. Collins, "Chapter 25 - Clinical Gait Analysis," in *Clinical Engineering*, A. Taktak, P. Ganney, D. Long, and P. White, Eds. Oxford: Academic Press, 2014, pp. 389–406.
- [30] E. Scheme and K. Englehart, "A comparison of classification based confidence metrics for use in the design of myoelectric control systems," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Milan, Aug. 2015, pp. 7278–7283, doi: 10.1109/EMBC.2015.7320072.
- [31] A. W. Wilson, Y. G. Losier, P. A. Parker, and D. F. Lovely, "A Bus-Based Smart Myoelectric Electrode/Amplifier—System Requirements," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 10, pp. 3290–3299, Oct. 2011, doi: 10.1109/TIM.2011.2159316.
- [32] E. J. Scheme and K. B. Englehart, "A flexible user interface for rapid prototyping of advanced real-time myoelectric control schemes," presented at the *Myoelectric Controls/Powered Prosthetics Symposium*, Fredericton, Canada, 2008.
- [33] Zhang, Yang, Qian, and Zhang, "Real-Time Surface EMG Pattern Recognition for Hand Gestures Based on an Artificial Neural Network," *Sensors*, vol. 19, no. 14, p. 3170, Jul. 2019, doi: 10.3390/s19143170.
- [34] S. M. Wurth and L. J. Hargrove, "A real-time comparison between direct control, sequential pattern recognition control and simultaneous pattern recognition control using a Fitts' law style assessment procedure," *J NeuroEngineering Rehabil*, vol. 11, no. 1, p. 91, 2014, doi: 10.1186/1743-0003-11-91.

- [35] M. Zia ur Rehman et al., "Multiday EMG-Based Classification of Hand Motions with Deep Learning Techniques," *Sensors*, vol. 18, no. 8, p. 2497, Aug. 2018, doi: 10.3390/s18082497.
- [36] E. Campbell, A. Phinyomark, and E. Scheme, "Linear Discriminant Analysis with Bayesian Risk Parameters for Myoelectric Control," in 2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Ottawa, ON, Canada, Nov. 2019, pp. 1–5, doi: 10.1109/GlobalSIP45357.2019.8969237.
- [37] L. J. Hargrove, K. Englehart, and B. Hudgins, "A Comparison of Surface and Intramuscular Myoelectric Signal Classification," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 5, pp. 847–853, May 2007, doi: 10.1109/TBME.2006.889192.
- [38] A. A. Adewuyi, L. J. Hargrove, and T. A. Kuiken, "Evaluating EMG Feature and Classifier Selection for Application to Partial-Hand Prosthesis Control," *Front. Neurobot.*, vol. 10, Oct. 2016, doi: 10.3389/fnbot.2016.00015.
- [39] S. Negi, Y. Kumar, and V. M. Mishra, "Feature extraction and classification for EMG signals using linear discriminant analysis," in 2016 2nd International Conference on Advances in Computing, Communication, & Automation (ICACCA) (Fall), Bareilly, India, Sep. 2016, pp. 1–6, doi: 10.1109/ICACCAF.2016.7748960.
- [40] A. Phinyomark, R. N. Khushaba, E. Ibáñez-Marcelo, A. Patania, E. Scheme, and G. Petri, "Navigating features: a topologically informed chart of electromyographic features space," *J. R. Soc. Interface.*, vol. 14, no. 137, p. 20170734, Dec. 2017, doi: 10.1098/rsif.2017.0734.
- [41] B. Lock, K. Englehart, and B. Hudgins, "Real-time myoelectric control in a virtual environment to relate usability vs. accuracy," presented at the MyoElectric Controls/Powered Prosthetics Symposium, Fredericton, Canada, 2005.

2022-01

Analyzing the impact of class transitions on the design of pattern recognition-based myoelectric control schemes

Raghu, Shriram Tallam Puranam

Elsevier

<https://doi.org/10.1016/j.bspc.2021.103134>

© This manuscript version is made available under the CC-BY-NC-ND 4.0 license <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Downloaded from UNB Scholar