IEEE ROBOTICS & AUTOMATION LETTERS
A PUBLICATION OF THE IEEE ROBOTICS & AUTOMATION SOCIETY

IEEE Xplore® Digital Library
http://ieeexplore.ieee.org/Xplore

Predicting Task Intent From Surface Electromyography Using Layered Hidden Markov Models

Yosef S. Razin[1], Kevin Pluckter[2], Jun Ueda[2], and Karen Feigh[1]

***Abstract*—Human–robot interaction faces the challenge of designing and modelling tightly coupled and effectively controlled human–machine systems. The objective of this research is to learn human operator's performance characteristics from surface electromyography measurements to predict their intentions during task operations. For the first time, a Layered Hidden Markov Model (LHMM) is successfully used with physiological data from co-contracting arm muscles to achieve accurate intent prediction. Furthermore, optimal model parameters and high-performing feature sets are identified and prediction accuracy at various time horizons calculated. The LHMM outperformed various other classification methods, including Naive Bayes and Support Vector Machine, ultimately achieving 82% accuracy in predicting the next 50 ms window of intent and maintaining 60% accuracy even after one second. These results hold the promise of improving robots' internal model of their human partners, which could increase the safety and productivity of human-robot teams in the factories of the future.**

***Index Terms*–Cognitive human-robot interaction, human-centered automation, physical human-robot interaction, gesture, posture and facial expressions**

I. Introduction

ON TODAY'S industrial assembly lines, a task's allocation is segregated by its operator: man or machine. With growing demand for customization, companies that have traditionally leveraged the speed, accuracy, and safety afforded by robotic manufacturing are reverting to manual operations [1]. More promising, however, is human-robot collaboration, which takes advantage of their combined strengths while mitigating each one's weaknesses.

Our research focuses on enabling the use of industrial robots with force-feedback control to provide the necessary strength to human operators for performing assembly line tasks that vary too much for automation, creating a complex, coupled system. While good models exist for controlling the system's robotic aspects, this is less true for its more unpredictable human components, and especially for any interaction effects that might occur. For example, industrial robots with force-feedback are prone to instabilities caused when human operators stiffen their arms while attempting to increase system impedance [2]. Therefore, the work herein described strives to characterize the physical human-robot interaction by leveraging physiological data from surface electromyography (sEMG) of select arm muscles directly recruited during robot operation. By measuring action potentials at the physical interface, we show that we can successfully predict the underlying intentions of the human operator.

While sEMG research has pushed toward activity recognition for more than 50 years, only recently have significant gains been achieved by using machine learning techniques [3]. Furthermore, these studies have been primarily within the orthotics and prosthetics community and each technique has been met with challenges [3]. Nevertheless, classification of sEMG signals by pattern-recognition methods has achieved accuracies ranging from $75-90\%$ [3]. Choice of features is known to have a more significant effect on classification accuracy than choice of machine learning technique. Dozens of EMG features have been identified, but selecting a subset or a lower-dimensional projection of those features can be challenging [3], [4]. Multiple studies have suggested specific feature sets for classification of sEMG [4]–[6], but these findings are often contradictory and may be influenced by experimental setup and design factors such as signal quality, sensor density, number of DFs, muscle synergies, and complexity of movement [5]. It has been shown that poor feature choices can cause classification error to increase by more than 50% [4], so features must be studied and selected carefully.

While nearly all other previous sEMG work has focused on classification of individual movements, this paper instead seeks to classify and predict current and future intentions. To do this successfully, higher-level intent must be learned from lower-level movement primitives, which themselves must be learned from observation. This structure naturally lends itself to a layered or stacked model such as a Layered Hidden Markov Model (LHMM), as originally proposed by Oliver *et al.* [7]. While LHMMs have been employed for activity recognition, they have been limited to using visual data from video decomposition [7]–[10]. However, such visual methods are not suited to the dynamic and cluttered environment of industrial robotics as well as wearable sEMG devices.

The goal of this work is to analyze classification and predictive performance of an LHMM using sEMG data and to identify proper features using the same data set as Moualeu

Manuscript received September 10, 2016; Revised November 26, 2016; Accepted January 10, 2017.

This paper was recommended for publication by Editor Dongheui Lee upon evaluation of the Associate Editor and Reviewers' comments. This work was supported by the National Science Foundation through the NSF NRI-Small Project Grant NSF IIS-1317718.

[1]School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA, USA. Email: yrazin@gatech.edu

[2]School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA, USA.

Digital Object Identifier (DOI): see top of this page.

Digital Object Identifier 10.1109/LRA.2017.2662741

et al. [11]. This research is novel in that it is the first to (1) successfully use an HMM-based technique to classify sEMG data for intent recognition, a method which outperforms its competitors, and (2) to successfully quantify predictions of future intent from sEMG signals. This work also attempts to clarify proper choice of feature set for this specific task by replicating previous findings in [4], [5]. In the future, we plan to use our results to dynamically update the gains of a haptic controller from predicted intent. Ultimately, integrating such capabilities will lead to more stable control and, thus, more effective human-robot collaboration.

II. Background

For modeling a human's interaction with a co-robot's haptic interface, multiple aspects must be considered, including: control, end-point stiffness, physiology, and an appropriate supervised learning technique for mapping hand movements to the controller [11]. This study focuses on the learning technique, building on previous work by Gallagher on suitable controllers and physiologic models [2], [12], [13].

A. Feature choice

This study looks to identify features for extraction that may provide highly predictive information. Both time and frequency domain feature extraction have been shown to provide better features than raw EMG data for determining arm muscle contractions, upper arm motions, and hand positions [4], [6], [14]. Specifically, root mean square (RMS) was shown to predict wrist torque using an SVM with high accuracy. Both RMS and 6th-order autoregressive coefficients along with other time-domain feature combinations showed success with an SVM for arm motion prediction [6]. A review of more than a dozen time domain and frequency domain features can be found in [4] and are fully listed in Table 1 of the Supplement.[1] However, a comparison of these techniques to predict intended motion in our specific task domain is still necessary for selecting an appropriate feature set.

[1]The Supplement is available at http://ieeexplore.ieee.org.

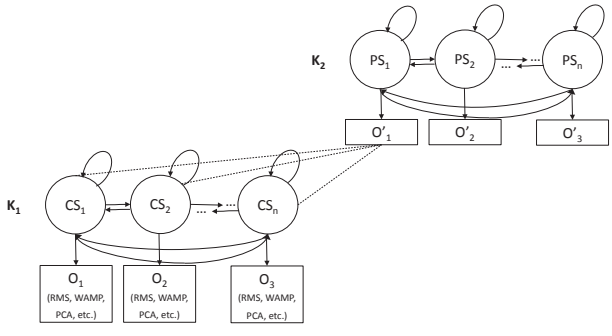

Fig. 1: Proposed Layered Hidden Markov Model with hidden prediction states PS, hidden classification states CS, and observations O composed of feature sets. We explore the appropriate number of nodes for each hidden layer K as well as choice of features. The lower layer K_1 outputs a classification that is used as the observation for the higher layer K_2.

B. Machine learning models

As mentioned above, a variety of methods have been employed to classify hand movements from sEMG signals for prosthetic control [3]. A highly accurate single-DF physical system was modeled ($\geq 90\%$) [15], a multi-DF linear classifier achieved high correlation coefficients ($\geq 79\%$) [16], and an Artificial Neural Network (ANN) achieved an even better average overall classification rate ($\geq 84\%$) [17]. Direct comparison of ANN against Twin SVM on the same data set showed that SVM had even higher recall (84% vs. 58%) and specificity (88% vs. 42%) [18]. Additionally, Omari and Gouhai [19] found varying success with Linear and Quadratic Discriminant Analysis (LDA & QDA) and k-Nearest Neighbor (KNN) using frequency domain features. Each of these techniques, though, has its own drawbacks. Controllers depend on highly accurate physiological models, and are often hard to scale to multiple DFs. Linear classifiers can be trained quickly but perform poorly with less separable data and non-linear noise, both issues in multi-DF sEMG signals. In contrast, ANNs and SVMs can handle non-linear classification boundaries but the former are slow to train and prone to over-fitting and the latter has significant real-time limitations and suffers from the curse of dimensionality.

Nonetheless, initial offline work attempted to classify operator stiffness and intended motion from sEMG data using SVMs [11], [20] but was unable to find a statistically significant improvement in performance as compared to a standard time-invariant linear-quadratic-Gaussian controller.

A Hidden Markov Model (HMM) may be more suited to time sequence analysis than an SVM, given the Markov assumption that the hidden states are only dependent upon one or more previous states and that observations are independent [21]. Earlier studies attempted to apply HMMs to two aspects of this particular problem. Gallagher *et al.* [13] used HMMs to classify stiffness levels using a small data sample and simplified, discretized observations to achieve 77% accuracy on unvalidated training data. Further work by Moualeu *et al.* [22] modelled three separate HMMs, one for each antagonistic muscle pair and a third for end-point velocity, and then used a decision HMM that endeavoured to guess which of the models best predicted the end-point stiffness. That work did not perform feature extraction and also modelled an HMM for each subject, making it poorly-generalizable.

To improve upon those results, LHMMs are proposed. Similar to Hierarchical Hidden Markov Models (HHMMs) [23], LHMMs are composed of multiple HMMs in which each layer provides a classification that is fed to a higher, more abstracted layer, as illustrated in Fig.1. Both models have been successfully used for activity recognition, achieving accuracies of $72-99\%$ on classification of visual data [7], [24]. While both HHMMs and LHMMs employ stacked HMMs, the former requires vertical transition probabilities and termination nodes, making it unsuitable for less linearly structured tasks. Moreover, LHMMs are trained layer-by-layer, building a modularized system that enables faster re-training [7].

III. Data collection and feature extraction

A. Data set

To enable comparison between feature extraction techniques, this research used the same data as Moualeu *et al.* [11], [22]. This data was collected from ten right-handed participants (7 male, 3 female; 21-31 y.) who manipulated a 1-DF haptic paddle device limited to a maximum generated force of 100 N at the handle. Participants were instructed to hold the device steady while command forces of randomly generated magnitude (0-40 N), direction (forward/backward), and duration (0.5-1.5 s) were applied at random intervals (2-20 s). The participants were required to apply a restoring force as needed to counter these disturbances and then relax their arm once the handle reached the desired position.

Data was collected with a multi-axis force/torque sensor (ATI Industrial Automation, Apex, NC), and a wireless EMG system (Cometa Srl, Milan, Italy) with true differential electrodes. One surface electrode was placed on each of four muscles that form two antagonistic pairs: the biceps brachii and triceps brachii in the upper arm and the flexor carpi ulnaris and extensor carpi ulnaris in the forearm.

The raw EMG signal was amplified by a gain of 1000, sampled at 2 KHz with a 16-bit analog-to-digital converter, and band-pass filtered between 10-500 Hz. Six state and dynamic features were measured at the end-effector: 1-DF state data S_{ee} = (position, velocity, acceleration) and 3-DF force data $F_3 = (F_x, F_y, F_z)$. For more information on the apparatus, sensors, and muscle selection, see [11], [20], [22].

B. Feature extraction

The studies previously cited performed minimal feature extraction on this data set [11], [20], [25]. In general, the main concern with sEMG signals is the high noise level, which can be considered white Gaussian noise when designing features. Phinyomark *et al.* [4] enumerated 16 methods of feature extraction from both time and frequency domains and compared their performances with KNN on identifying hand position. All of the these extraction techniques use sliding windows, and 15 of them were implemented here, some with the MEC toolbox [26]; see Table 1 in the Supplement. Englehart *et al.* [5] found that Principal Component Analysis (PCA) following feature selection consistently performed better than feature selection alone. Therefore, tests using principal components of the sEMG data were also performed in order to validate their findings.

IV. Layered Hidden Markov Model

Layered Hidden Markov Models [7] can be formally described as stacked or nested HMMs consisting of $N+1$ layers, where the bottommost layer is composed of a time series of observations $o_{1:T}$ within the emission vector space O. Given $o_i \in O$, the N^{th} hidden layer HMM is a multi-class mapping $f_N : O \rightarrow S^N$, where S^N are the hidden state labels for the N^{th} HMM. At the next level, $N-1$, the outputs of the previous HMM, S^N are now taken as the new inputs, so that $f_{N-1} : S^N \rightarrow S^{N-1}$.

Unlike previous studies [13], [22], where the end-point stiffness served as the hidden variable, we now choose to model each intention with its own HMM composed of hidden classification states (*CS* in Fig.1) and perform generative classification. A higher level HMM is then used to determine the next predicted state (*PS*) of intention. Prediction is performed by repeatedly sampling the trained HMM sequentially and taking the most likely posterior observations. Thus, we repeatedly sample the probability distribution $P(o_{n+1}|o_n, \lambda)$ from $n \in [t, \tau]$ where o_n and o_{n+1} are the current and next observation respectively, λ is the trained HMM model, and $[t, \tau]$ represent the points from the current sample's index t out to prediction horizon τ. Each HMM was trained with the Baum-Welch algorithm and validated individually, the latter being one of the benefits to the parallel and modular structure of the LHMM as a whole [7]. The full LHMM was then tested on previously unseen data from all ten users. The HMM framework from Murphy's PMTK-3 toolbox [27] was used for modelling all HMM's.

A. Classification parameters

For both the prediction and classification layers of the LHMM, parameters such as the number of hidden nodes and the set of features which yielded the lowest error rate were identified. Since decreasing the number of features makes computation more tractable, feature selection and PCA were both tested and compared, to validate Englehart's findings [5]. For feature selection, the classification HMMs were trained with two dozen feature subsets, chosen in meaningful groups and specified by feature set number, per Table 2 of the Supplement. PCA was performed by simply increasing the number of principal components (PCs) kept for analysis (1–20) for each EMG channel. Four categories of features were tested: feature sets (FS), PCs of the EMG components only (E), PCs of all components (F), and the EMG PCs with the six state and dynamics features (6+E).

The data from all ten trials was first manually labelled to allow error measurement and then broken down into run lengths of the three intention states: *Ready, Move* and *Hold*. Data from all ten trials was combined into a single training set for classification. A full bi-variate analysis was run with 10-fold validation looking at the effect on classification error of all feature sets, as detailed in Fig. 1 of the Supplement.

Similarly, the prediction layer was tested using 10-fold validation to identify the number of hidden nodes that yielded the lowest error rate in prediction over time (Fig. 2). However, instead of dividing up the labelled data by run length, as above, it was divided into 50-ms non-overlapping windows, each labelled with the mode of that window. Since the sequential order of the data is now important, each trial was tested separately and the error reported now is the mean prediction error.

The LHMM classifier was also compared against six other classification algorithms: Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Naive Bayes (NB), Decision Tree (DT), KNN, and SVM. For each of the seven classifiers, 13 feature subsets from feature selection (indicated

in Table 2 of the Supplement) as well as the first four and ten PCs of E, F, and 6+E were run, yielding 133 total combinations. Each individual combination was run with 5-fold validation, and the results of the best-case performance for each technique are given in Table I.

B. Intent prediction

The full LHMM, illustrated in Fig. 1, was composed of the two previously trained layers and tested on heretofore unseen 50-ms-windowed data. Window size was chosen based on sampling frequency and applicability to real-time control. Given a window of observation data, the LHMM first classified it. The top layer HMM was then sampled 300 times for a given prediction horizon and the mode of those sequences taken, similar to the *maxbelief* model used in [7]. To maintain separate training and testing sets, 10-fold validation was accomplished by selecting ten windows at random in each trial per prediction horizon and then averaging across trials for a given horizon to get the mean prediction error.

C. Evaluation

For both the LHMMs individual layers and their combination, the main metric was error, calculated as $e = \frac{1}{N}\sum_{n=1}^{N}(\hat{y}_n \neq y_n)$, where $\hat{y}_n$ is the predicted class of observation n out of N total observations, and y_n is the true value. Thus, error signifies the percentage of misclassifications or false predictions. Further metrics, such as precision, recall, and F1 scores (the harmonic mean of precision and recall) were also taken on the classification layer of the LHMM as well as the other classifiers. The F1 score is the best metric for classifier comparison here given the unbalanced classes of equal import, as F1 signifies the minimum combination of both Type I and Type II errors. Throughout this work, accuracy is defined as $1 - e$.

V. Results & Discussion

A. Parameter estimation

Analysis of feature selection techniques confirmed that, as expected [3], in the classification layer, the choice of features has a much stronger effect on error than the number of nodes. A two-way ANOVA for number of nodes against features yielded $\eta_p^2 = .96$ for number of PCs and $\eta_p^2 = .99$ for choice of feature set. Given that a majority of the variance is explained by feature choice and not by node number, three hidden nodes were used in the LHMMs classification layer.

TABLE I: Best-case performance for each algorithm over all feature sets, ranked by error rate (*e.g.* the best-case SVM ran with FS #9 and ranked 29^{th} out of 133 algorithm/feature set combinations). Feature set #s are defined in Table 2 of the Supplement

Rank	Method	Features	Error	Precision	Recall	F1
1	LHMM	FS #10	0.159	0.762	0.780	0.735
10	DT	E (10-PC)	0.229	0.686	0.687	0.685
11	NB	FS #13	0.242	0.681	0.654	0.659
15	QDA	FS #9	0.275	0.660	0.623	0.621
29	SVM	FS #9	0.332	0.611	0.518	0.453
39	LDA	FS #9	0.347	0.522	0.512	0.484
50	KNN	FS #23	0.363	0.532	0.536	0.532

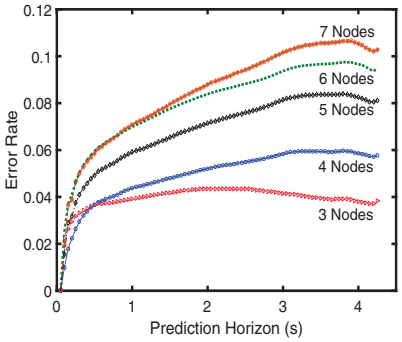

Fig. 2: Number of hidden nodes vs. error in future predictions. For the upper-level prediction HMM, four nodes performed best during the first 0.4 s ($e \leq 3.8\%$), while three nodes performed better ($e \leq 4.2\%$) beyond that horizon.

The number of nodes in the prediction layer does have a significant effect, as can be seen in Fig. 2. The general trend shows that the error increases with number of nodes, likely due to over-fitting. Within the first 0.4 s, four nodes performed best; beyond that three nodes performed better. Thus, both 3- and 4-node models performed well enough that each was implemented in the full LHMM, to enabled further comparison.

B. Individual layer error

The highest average validated accuracy rate achieved by any feature choice during classification was 84% (F1= 0.74). This falls within the expected range of $75-90\%$ for sEMG data [3], [18] as well as the $70-90\%$ range for activity recognition with LHMM [7], [8]. As can be seen in Table I, LHMM outperformed all other methods both in terms of lowest error and highest F1 score. In fact, the top 9 of 133 combinations of features and learning algorithms were with the LHMM classifier, six using FS features and three using E+6 features.

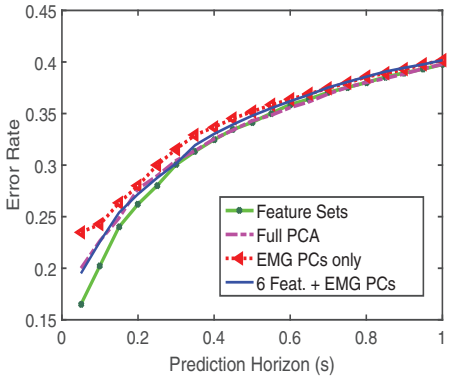

Fig. 3: Best-case performance for each feature reduction method over the next second in the full 3-node LHMM. Feature set #7 performed better or as well as any other performer for all horizons during the first second with EMG PCs performing worst.

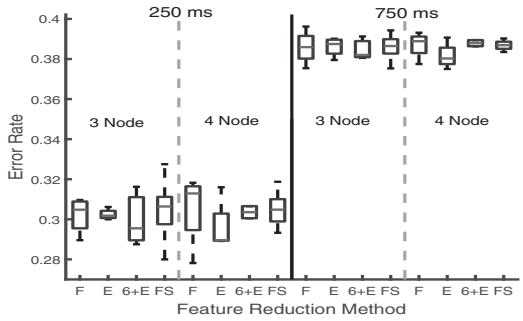

Fig. 4: Three- and four-node LHMM for 250-ms- and 750-ms-prediction errors. F is the PCs of the full data set, E the PCs of the EMG features only, 6+E the six state and dynamic features (S_{EE}, F) with the EMG features. The $4^{th}, 5^{th}, 10^{th}$, and 13^{th} PC sets are tested for all three PC methods. FS is feature selection and represents all feature sets with normalized EMG data.

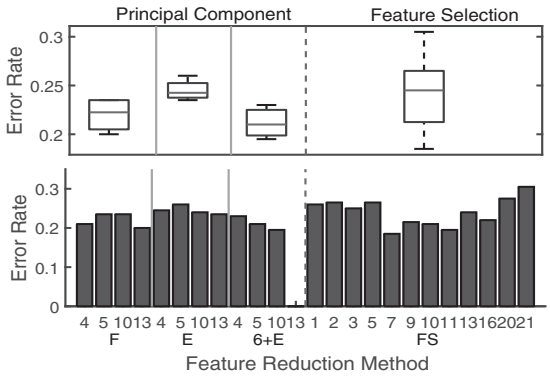

Fig. 5: One-step look ahead performance of all sets. Feature reduction technique performance (above) and feature groups within techniques (below) compared in predicting the next 50 ms window using the 3-node model. The 6+E method with 13 PCs failed to compute and is therefore absent.

Six of the seven learning techniques did best in combination with FS in contradistinction to Englehart's findings [5].

During prediction, error grew with increasing horizon as expected. All models were able to achieve under 1% error (given perfect classification) for the first few frames (100-150 ms), but that rapidly increased with horizon length. The lowest overall error rate was achieved by the 3-node model, where no errors exceeded 4.2% even beyond 4 s. Thus, the limiting factor for LHMMs predictive ability was classification accuracy, and future improvement efforts would be best spent there.

C. Feature choice

The single-layer results above were used to help guide selection from 178 features[2] for further testing on the full LHMM. The sets of PCs that returned the minimal error were the first 4, 5, and 13. Given the fairly smooth relationship between PCs and error rate, a fourth set of PCs (first 10) was also run to get a more complete picture.

In general, all feature selections that included end-effector states and forces had lower error relative to those that did not include them. While this observation was not used to eliminate features from testing, it led us to consider a combination feature set of these 6 features with EMG PCs for comparing the predictive ability gained from considering the EMG data (6+E mentioned above).

Finally, while [3] suggested that normalizing the EMG signals to maximum voluntary contraction might reduce amplitude cancellation, we did not find this to be the case. Only the normalized EMG data was kept to remain consistent with [11], [22].

D. Intent prediction with two-layer LHMM

In general, using the best feature selections yielded fairly strong and similar results regardless of selection method. Fig. 3 shows the best-case results for each feature type, after running though the full two-layer LHMM with each layer containing 3 nodes. Every method yielded an error rate below 25% for the next 50-ms frame and after that the error grew slowly over a one second horizon to 40%. The main difference between feature selection methods on final error rate is near-horizon performance, where the best performing feature set (#7) performed slightly but noticeably better ($\approx 3\%$) than the PC-based methods. Most of the error in the full LHMM is likely due to misclassification and not prediction, as the prediction error rate tended to be an order of magnitude smaller than the classification error rate.

Testing of the full LHMM allowed us to compare 3- vs 4-node models for short- and long-term performance. Fig. 4 illustrates with which feature reduction technique three nodes worked better than four. However, against expectations, no clear preference emerged for either three or four nodes. In fact, prediction error across node numbers and feature sets differed only by, at most, a few percent ($\leq 4\%$).

Fig. 5 compares how each feature reduction and selection technique performed during one-step look-ahead prediction using the 3-node model. On the whole, F and 6+E performed better than methods E and FS. The generally improved performance of 6+E may lend support to the findings of Englehart *et al.* [5]. On the other hand, FS had a larger variance than the PC-based models, which is actually promising, as it means some of the individual sets performed especially well, in fact better than any of those from the PC–based methods.

The standout feature set, FS #7, consists of the normalized EMG data and co-contraction data only and achieves 82% accuracy at 50 ms, falling to 60% accuracy after 1 s. This is surprising considering that this feature set does not include the 6 state and force features, which had the best classifier performance (see above). The next best feature set (#11) is comprised of the 6 state/force features and the 3 features suggested by [4] (HEMG, WAMP, MMNF, as defined in the Supplement) extracted from the normalized EMG data, lending support to their hypothesis that these form one of the best feature combinations. In contrast, the other features suggested

[2] 18 raw signals including 4 raw and 4 normalized EMG channels, with 20 features extracted per EMG channel, as described in Table 1 of the Supplement

by Phinyomark *et al.* [4], such as MAV, WL, and RMS, were included in the full feature set (#14) and incorporated in the PC-based sets but yielded unsatisfactory results.

The decreasing performance with lengthening horizon of the LHMM stands in contrast to past findings by Patel *et al.* [24]. They found that both the upper and lower level HMMs' accuracies improved from "short" to "long" time horizon, going from $63 \rightarrow 74\%$ and $72 \rightarrow 86\%$, respectively. While we only took horizon length into account in the upper layer, our accuracy fell from 82% to 60% over one second. Thus, there is evidently a trend reversal. However, it is unclear from their work what periods of time are indicated by "short" or "long". Furthermore, even though they showed that using an SVM to do the lower level classification may be beneficial, these gains are qualified by their lack of robustness in new situations and with missing data [24]. While new situations are unlikely in an industrial setting, corrupted data is common in sEMG due to high noise, amplitude cancellation, and cross-talk [3].

VI. Conclusion

Working toward accurate intent prediction using sEMG signals from physical human-robot interaction, this research found that both PCA and feature selection of sEMG features produce fairly accurate predictions. This partially confirms findings from [4] and [5]. It also showed that a 3-node classification layer combined with either a 3- or 4-node prediction layer produces the lowest prediction error. This work demonstrated that in the context of sEMG data from a haptic device, LHMM classification outperformed other machine learning techniques. This work furthered the promise of earlier research by novelly learning actions and intentions from sEMG data using an LHMM and making successful predictions with 82% accuracy for the next 50 ms and 60% accuracy over the next second. Future work will examine other possible combinations of the 178 extracted features, better feature reduction methods, and improvement of the classification layer. Lastly, this work will be applied to a physical system to test the LHMMs ability to update a controller in real-time to produce better stability at the interface of man and machine.

References

[1] S. Gibbs, "Mercedes-Benz Swaps Robots for People on its Assembly Lines," *The Guardian*, February 2016. [Online]. Available: https://www.theguardian.com/technology/2016/feb/26/mercedes-benz-robots-people-assembly-lines

[2] W. J. Gallagher, "Modeling of Operator Action for Intelligent Control of Haptic Human-Robot Interfaces," Ph.D. dissertation, Georgia Institute of Technology, December 2013.

[3] M. Ison and P. Artemiadis, "The Role of Muscle Synergies in Myoelectric Control: Trends and Challenges for Simultaneous Multifunction Control," *J. of Neural Eng.*, vol. 11, no. 5, pp. 051 001–051 022, September 2014.

[4] A. Phinyomark, C. Limsakul, and P. Phukpattaranont, "A Novel Feature Extraction for Robust EMG Pattern Recognition," *J. of Computing*, vol. 1, pp. 71–80, December 2009.

[5] K. Englehart, B. Hudgins, P. A. Parker, and M. Stevenson, "Classification of the Myoelectric Signal using Time-Frequency Based Representations," *Med. Eng. & Phys.*, vol. 21, no. 6, pp. 431–438, July 1999.

[6] M. A. Oskoei and H. Hu, "Support Vector Machine-Based Classification Scheme for Myoelectric Control Applied to Upper Limb," *IEEE Trans. Biom. Eng.*, vol. 55, no. 8, pp. 1956–1965, August 2008.

[7] N. Oliver, A. Garg, and E. Horvitz, "Layered Representations for Learning and Inferring Office Activity from Multiple Sensory Channels," *Comp. Vis. & Image Understanding*, vol. 96, no. 2, pp. 163–180, August 2004.

[8] D. Aarno and D. Kragic, "Layered HMM for Motion Intention Recognition," in *2006 IEEE/RSJ Int. Conf. Intell. Robots and Sys.* IEEE, October 2006, pp. 5130–5135.

[9] T. L. van Kasteren, G. Englebienne, and B. J. Kröse, "Hierarchical Activity Recognition using Automatically Clustered Actions," in *Ambient Intell.* Springer, November 2011, pp. 82–91.

[10] S. Luhr, H. H. Bui, S. Venkatesh, and G. A. West, "Recognition of Human Activity Through Hierarchical Stochastic Learning," in *PerCom 2003: Proc. 1st IEEE Int. Conf. Pervasive Comp. and Commun.* IEEE, March 2003, pp. 416–422.

[11] A. Moualeu, W. Gallagher, and J. Ueda, "Support Vector Machine Classification of Muscle Co-contraction to Improve Physical Human-Robot Interaction," in *2014 IEEE/RSJ Int. Conf. Intell. Robots and Sys.* IEEE, September 2014, pp. 2154–2159.

[12] W. Gallagher, M. Ding, and J. Ueda, "Relaxed Individual Control of Skeletal Muscle Forces via Physical Human–Robot Interaction," *Multibody Sys. Dyn.*, vol. 30, no. 1, pp. 77–99, June 2013.

[13] W. Gallagher, D. Gao, and J. Ueda, "Measurement of Muscle Stiffness to Improve Stability of Haptic Human-Robot Interfaces," in *ASME 2012 5th Annu. Dyn. Sys. and Contr. Conf. joint with the JSME 2012 11th Motion and Vibr. Conf.* American Society of Mechanical Engineers, October 2012, pp. 493–502.

[14] Z. O. Khokhar, Z. G. Xiao, C. Menon, *et al.*, "Surface EMG Pattern Recognition for Real-Time Control of a Wrist Exoskeleton," *Biomed. Eng. Online*, vol. 9, no. 41, August 2010.

[15] K. A. Wheeler, D. K. Kumar, and H. Shimada, "An Accurate Bicep Muscle Model with sEMG and Muscle Force Outputs," *J. Med. Biol. Eng.*, vol. 30, no. 6, pp. 393–398, December 2010.

[16] P. K. Artemiadis and K. J. Kyriakopoulos, "EMG-based control of a robot arm using low-dimensional embeddings," *IEEE Trans. Robot.*, vol. 26, no. 2, pp. 393–398, April 2010.

[17] M. R. Ahsan, M. I. Ibrahimy, and O. O. Khalifa, "Electromygraphy (EMG) Signal Based Hand Gesture Recognition using Artificial Neural Network (ANN)," in *2011 4th Int. Conf. Mechatronics (ICOM).* IEEE, May 2011, pp. 1–6.

[18] G. R. Naik, D. K. Kumar, and Jayadeva, "Twin SVM for Gesture Classification using the Surface Electromyogram," *IEEE Trans. Inf. Tech. Biomed.*, vol. 14, no. 2, pp. 301–308, March 2010.

[19] F. Al Omari and L. Guohai, "Analysis of Extracted Forearm sEMG Signal using LDA, QDA, K-NN Classification Algorithms," *J. Open Automat. and Contr. Sys.*, vol. 6, pp. 108–116, June 2014.

[20] A. Moualeu and J. Ueda, "Haptic Control in Physical Human-Robot Interaction Based on Support Vector Machine Classification of Muscle Activity: A Preliminary Study," in *ASME 2014 Dynamic Systems and Control Conference*. American Society of Mechanical Engineers, October 2014.

[21] L. R. Rabiner and B.-H. Juang, "An Introduction to Hidden Markov Models," *IEEE ASSP Magazine*, vol. 3, no. 1, pp. 4–16, January 1986.

[22] A. Moualeu and J. Ueda, "A Model for Operator Endpoint Stiffness Prediction during Physical Human-Robot Interaction," *J. of Human-Robot Interact.*, vol. 4, no. 3, pp. 170–193, December 2015.

[23] S. Fine, Y. Singer, and N. Tishby, "The Hierarchical Hidden Markov Model: Analysis and Applications," *Mach. Learning*, vol. 32, no. 1, pp. 41–62, July 1998.

[24] M. Patel, J. V. Miro, D. Kragic, C. H. Ek, and G. Dissanayake, "Learning Object Grasping and Manipulation Activities using Hierarchical HMMs," *Autonomous Robots*, vol. 37, no. 3, pp. 317–331, June 2014.

[25] W. Gallagher, D. Gao, and J. Ueda, "Improved Stability of Haptic Human–Robot Interfaces using Measurement of Human Arm Stiffness," *Adv. Robotics*, vol. 28, no. 13, pp. 869–882, 2014.

[26] A. D. Chan and G. C. Green, "Myoelectric Control Development Toolbox," in *Proc. of 30th Conf. Canadian Med. & Bio. Eng. Soc.*, vol. 1, August 2007, p. M0100.

[27] K. Murphy, "PMTK3: Probabilistic Modeling Toolkit for MATLAB/Octave, Version 3," (2016). [Online]. Available: https://github.com/probml/pmtk3, Accessed on: Nov. 23, 2016.