From "MEC '08 Measuring Success in Upper Limb Prosthetics," Proceedings of the 2008 MyoElectric Controls/Powered Prosthetics Symposium, held in Fredericton, New Brunswick, Canada, August 13–15, 2008.

ADAPTIVE PATTERN RECOGNITION TO ENSURE CLINICAL VIABILITY OVER TIME

Jonathon W. Sensinger[1], Blair A. Lock[2], and Todd A. Kuiken[2,3]

1) *Contact* jsensinger@gmail.com, People-Inspired Prostheses Laboratory,
Center for Biomedical and Robotic Technology, *Mahidol University*, Bangkok, Thailand

2) Neural Engineering Center for Artificial Limbs, *Rehabilitation Institute of Chicago*, Chicago, IL

3) Department of Physical Medicine and Rehabilitation, *Northwestern University*, Evanston, IL

ABSTRACT

Pattern Recognition is a useful tool for deciphering movement intent from myoelectric signals. In order to be clinically viable over time, recognition paradigms must be capable of adapting with the user. Most existing paradigms are static, although two forms of adaptation have received limited attention: *Supervised* adaptation achieves high accuracy, since the intended class is known, but at the cost of repeated cumbersome training sessions. *Unsupervised* adaptation attempts to achieve high accuracy without explicitly being told the intended class, thus achieving adaptation that is invisible to the user at the cost of reduced accuracy. This paper reports a novel adaptive experiment on eight subjects that allowed a post-hoc comparison of four supervised and three unsupervised adaptation paradigms. All supervised adaptation paradigms reduced error over time by at least 23%. Most unsupervised adaptation paradigms failed to achieve statistically significant reductions in error due to the uncertainty of the correct class. One method that selected high-confidence samples showed the most practical potential, although other methods warrant future investigation outside of a laboratory setting. The ability to provide supervised adaptation should be incorporated into any clinically viable pattern recognition controller, and unsupervised adaptation should receive renewed interest in order to provide invisible adaptation.

INTRODUCTION

Myoelectric prosthesis control may be compared with recognition of a person in a photograph. If you are only concerned with the height of the person, then the silhouette of the person may be sufficient to calculate their height. Likewise, the average myoelectric signal amplitude is sufficient to calculate the speed or position of the prosthesis if some control strategy such as co-contraction will determine which joint will move.

If, on the other hand, you are asked to identify the person in the photograph, features in addition to the silhouette are required. These features may come from a variety of sources, including color (hair, eyes, skin, etc…), shape (round face, chiseled face), etc… By combining all of these features, you have a better probability of successfully identifying the person. Likewise, it is fairly straightforward to identify a desired grasp pattern or movement encoded in the electrical image produced by the muscle ensemble by looking at numerous features of myoelectric signals. Such a concept is termed pattern recognition and has been applied to myoelectric control since the 1970's [1]. Substantial progress towards a clinically viable system made in the 1990's [2] combined with the introduction of powerful microcontrollers has allowed companies to begin designing a new generation of prostheses that are capable of recognizing these myoelectric patterns, allowing for more functional and lifelike movement of multifunctional prostheses.

The above illustrations ignore an important consideration: the possibility of temporal change between training and testing. Suppose between the time you were trained to recognize a person's picture and the time you were tested, he or she cut their hair, got sunburned, and lost twenty pounds. Your ability to correctly identify this person (class) within a group of similar persons (classes) would depend on how well you had been trained, with four possible scenarios:

a) **Singular Training:** If you had only been shown the original and final image of the person, identification would be difficult.
b) **Robust features** [3]**:** If you had based your decision primarily on robust features, such as scars or tattoos, rather than on features that may frequently change, then the effect of temporal changes would not substantially affect your decision.
c) **Robust training** [4]: If during your initial training you had been shown many different images of the person (i.e., complete wardrobe, several hair cuts, pictures from different angles, etc…), it is likely that you would have recognized the person despite the changes.

This method assumes that all change falls within the permutation range of the robust training session. It also requires lengthy permutations of all of the variables that could change (i.e., losing weight and wearing heavy clothing; gaining weight and wearing light clothing, etc…)

d) **Adaptive Training** [5, 6]: If you had only been shown one image the first day of training, but you had also been shown periodic updates, your mental image of the person would adapt over time, allowing you to recognize gradual changes such as weight loss, and possibly even drastic changes if they only happened one at a time. If the identity of the person in the updated picture is provided, the adaptation is termed *supervised [6]*. If the identity of the person in the updated picture is not provided, the adaptation is termed *unsupervised [5]*.

Pattern changes over time are an unavoidable reality both in pictures and myoelectric signals. Myoelectric pattern changes can be caused by a variety of factors, including electrode conductivity changes (perspiration, humidity), electrophysiological changes (muscle fatigue, atrophy, or hypertrophy), spatial changes (electrode shifting or tissue fluid fluctuations), and user changes (adaptation or contraction intensity changes). Despite these temporal changes, the majority of reported classifiers rely on a single training session. Some studies only use signal amplitude, which is not temporally robust, in conjunction with a complex classifier [5]. Others use a slightly more robust ensemble of features with a simpler classifier [3]. Both constructions achieve high classification accuracy [7, 8], even when testing subjects with an amputation [8, 9], since they both train and test within the same session. These results may not be clinically viable, however, because they do not show robustness over time.

Other techniques seek to use robust training. Although this technique has produced encouraging results when applied to a single element of change, such as electrode shifting [4], incorporating all of the possible fluctuations into the data set requires too many permutations to be clinically viable.

The Neural Engineering Center for Artificial Limbs performs research on subjects with targeted muscle reinnervation (TMR): a novel surgical technique that provides the residual muscles of amputees with similar signal content to the normal muscles of able-bodied subjects [9-12]. Because we see the same subjects several times each year, our laboratory has turned its attention to finding a long-term clinically viable pattern recognition approach. Towards this end, we have begun a systematic analysis of the fourth solution: how to create a clinically viable classifier that gradually adapts to the user's changing patterns over time in a manner such that the user is unaware that the system is even adapting. This paper reports our progress in this critical area of clinically viable pattern recognition control.

METHODS

Five able-bodied subjects and three TMR amputee subjects (two Shoulder Disarticulation level and one Transhumeral level) participated in the study. All procedures were performed with informed consent and approved by the Northwestern University Institutional Review Board. Signals were pre-amplified and filtered using commercially available myoelectric amplifiers[1] and recorded with a custom-built 16-bit EMG amplification and acquisition system at a sampling rate of 1 kHz. Electrodes were placed equidistantly around the circumference of the proximal third of the arm with a longitudinal orientation for able-bodied subjects. Electrodes were placed at predetermined locations on subjects with TMR that had yielded the best classification accuracy using a high-density electrode array [12]. Four features[2] were extracted from each of twelve electrodes every 30 ms in 150 ms overlapped bins. A large number of classes (eleven[3]) were tested using an LDA classifier. A large number of classes were chosen to make classification difficult. Increased difficulty in turn should amplify any robustness problems, and provide room for adaptation. Custom software[4] was used to process data and

---

[1] Liberating Technologies, Inc. BE328 Remote AC electrodes, 30 Hz – 420 Hz -3dB bandpass filter

[2] Features included mean absolute value, # of zero-crossings, waveform length, and #of slope sign changes

[3] For amputees, the 11 classes included elbow flexion/extension, forearm pronation/supination, wrist flexion/extension, hand open, 3 self-selected grasp patterns, and no movement. For able-bodied subjects, the 11 classes included forearm pronation/supination, wrist flexion/extension, hand open, 5 grasp patterns (3-jaw chuck, lateral key, fine pinch, trigger, power), and no movement.

[4] Acquisition & Configuration Environment (ACE), a myoelectric control software program developed by the University of New Brunswick

From "MEC '08 Measuring Success in Upper Limb Prosthetics," Proceedings of the 2008 MyoElectric Controls/Powered Prosthetics Symposium, held in Fredericton, New Brunswick, Canada, August 13–15, 2008.

provide a graphical user interface. Subjects alternated between training and testing trials for a total of ten pairs of training and testing data. Trials contained two repetitions of each class held for 3-4 seconds. Classifiers calculated from each training trial provided real time visual feedback for the subsequent (paired) testing trial. The experiment lasted approximately two hours, including a sum of one hour's worth of muscle contractions.

Because testing trials provided real-time feedback between each training dataset used to calculate a classifier, the static classifiers obtained over the two-hour long session may be thought of as static snapshots of an adaptive classifier. **This strategy presented a novel protocol for assessing adaptive classifiers that allowed rigorous post-hoc comparison of different adaptation paradigms while preserving the dynamic qualities of real-time adaptation.**

Post-Hoc Adaptation Comparison. Adapted classifiers were calculated post-hoc from data sets that combined data from the original classifier (of the first three classifiers, this was the classifier which had the lowest classification error) with selected samples from the testing trials. Real-time testing samples were selected based on one of two qualities, including the confidence[5] of the decision and how consistently a given class was selected. Confidence is defined as $C = \sum_{k=1}^{K} p_k \ln(p_k)$, where p_k is the probability of class k and K is the number of classes to be considered.

Two types of adaptive classifiers were compared. The first was:

Supervised Adaptation Paradigms (correct class provided):

- Supervised High confidence *(SH)*: Add samples with high confidence, with known intended class.
- Supervised Low confidence *(SL)*: Add samples with low confidence, correcting the class of low-confidence decisions.
- Supervised High/Low confidence *(SHL)*: Add samples with high or low confidence, correcting the class of low-confidence decisions.
- Supervised All *(SA)*: Add all samples, with known intended class.

The second was:

Unsupervised Adaptation Paradigms (algorithm must guess correct class):

- Unsupervised High confidence *(UH)*: Add samples with high confidence. Assume the predicted class is correct [5].
- Unsupervised Low confidence *(UL)*: Add samples with low confidence. Guess the correct class from 2nd choice guess and surrounding samples.
- Unsupervised Blip *(UB)*: Add samples that blipped to another class and then settled back to the same class. Assume the class before/after the blip is correct [13].

Data Processing. Only data following the subject's response to the visual prompt was analyzed. Tuning parameters for each adaptation paradigm were optimized based on pilot data from three other subjects. The original classifier was selected as the baseline classifier to ensure a reasonable starting trial. Adaptation paradigms added successive groups of samples to this data set for remaining trials. Relative error[6] reductions between the non-adapting classifier and the adaptive classifiers were averaged across the remaining 6-8 trials for each subject. Error from

[5] Entropy was the actual metric used, but for the purposes of this paper I have used the term confidence and adjusted the equation to better convey the meaning

[6] Relative reductions in error, rather than absolute reductions in error, are reported in the results, because relative errors retain more information content when averaged across multiple trials and multiple subjects .Information content was assessed by measuring Kurtosis, the fourth standardized moment of a distribution. Kurtosis was measured for each subject and found to be larger across subjects for relative error (2.5) than for absolute error (2.2), p = 0.06

From “MEC '08 Measuring Success in Upper Limb Prosthetics,” Proceedings of the 2008 MyoElectric Controls/Powered Prosthetics Symposium, held in Fredericton, New Brunswick, Canada, August 13–15, 2008.

the non-adapting and adaptive classifiers was fit across trials using a linear least squares regression. The statistical power of the slope determined if subject error changed over time. Minimum error was calculated by testing and training on the *same* real-time data. Any errors reported in minimum error are simply due to overlap in feature space, and may only be solved by better feature sets, more sophisticated classifiers, or better distinction by the user.

RESULTS

Adaptation paradigms frequently tagged samples to add to the baseline classifier (Figure 1). The top of this figure shows the prompted class as circles and original predictions as asterisks. The confidence of each decision is shown in the middle of the figure with a gray background. Post-hoc adaptation strategies tagged some samples to add to the training set. Adaptation strategies either used the prompted class (supervised strategies), kept the same predicted class (*UH*), or suggested a new predicted class (*UL*, *UB*) if they thought the original classifier had made an error. All of the unsupervised adaptation paradigms incorrectly tagged some samples. Non-adapting error grew larger over time for many subjects (Figure 2), and many adaptation strategies were able to reduce the average error and prevent error from increasing over time.

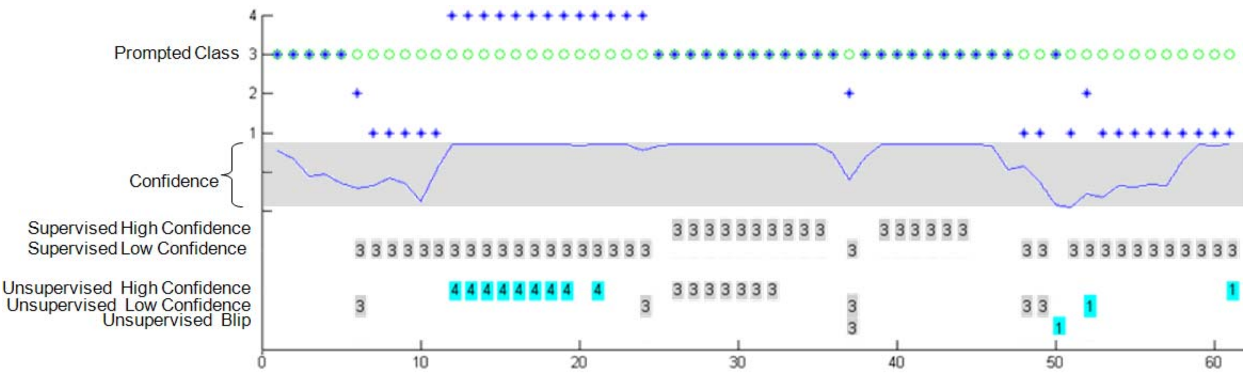

Figure 1: Composite snapshot of, from top to bottom: real-time test (circles/*), confidence of decisions (line), and samples tagged in post-hoc analysis for inclusion in the adaptive data set (numbers). Out of 11 classes, only 1 class was prompted (shown as green circles) during this snapshot and 4 classes were predicted (shown with blue *) during the real-time test. Post-hoc adaptation paradigms selected different samples to add to the adapted data set, based on criteria specific to each adaptation paradigm. Sometimes unsupervised adaptation strategies suggested the inclusion of a sample but incorrectly changed the class: these incorrectly tagged samples are highlighted in cyan.

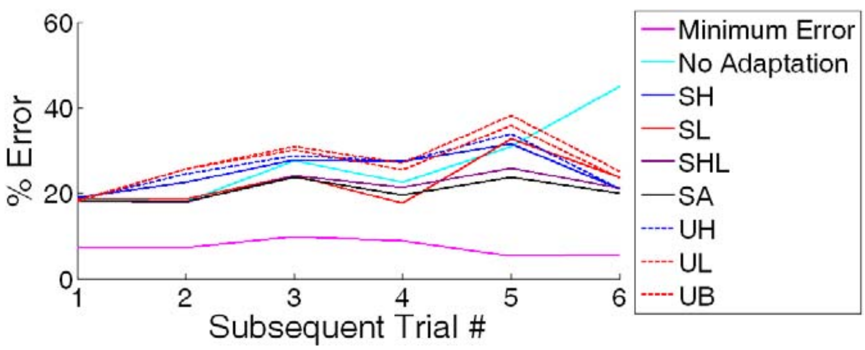

Figure 2: 2 Example of error over time for a typical subject. Subject TMR2’s non-adaptive classifier produced increasing error over time.

Subjects typically had large errors when the initial non-adapting classifier was used on the data throughout the entire two-hour session. Many adaptation strategies successfully reduced this error. The average non-adapting error across subjects was 25%. All supervised adaptation paradigms reduced the error of the classifier ($p<.03$). Supervised High-confidence adaptation *(SH)* had a 27% relative reduction in error, Supervised Low-confidence adaptation *(SL)* had a 23% relative reduction in error, Supervised High/Low-confidence adaptation *(SHL)* had a 29% relative reduction in error, and adding all of the samples *(SA)* had a 30% relative reduction in error. *SA* provided more reduction in error than its more selective

From "MEC '08 Measuring Success in Upper Limb Prosthetics," Proceedings of the 2008 MyoElectric Controls/Powered Prosthetics Symposium, held in Fredericton, New Brunswick, Canada, August 13–15, 2008.

alternatives (SH, SL, SHL) ($p<0.02$). None of the selective adaptation paradigms exhibited statistically significant differences from each other ($p>0.08$).

Unsupervised adaptation paradigms did not reduce error as much as supervised adaptation paradigms. The Unsupervised High-confidence *(UH)* paradigm showed the most relative reduction in error (21%), which bordered on statistical significance across subjects ($p=0.053$). Neither Unsupervised Low-confidence *(UL)* nor Unsupervised Blips *(UB)* had statistical significance ($p>0.07$) and had lower relative reductions in error (*UL*: 19%, *UB*: 16%). These results are likely due to the fact that adaptation paradigms that had to recalculate the predicted class (*UL* & *UB*) were frequently unable to correctly guess the correct class: 35% of the samples added by the *UL* paradigm and 54% of the samples added by the *UB* paradigm incorrectly reassigned the class. A small percentage (9%) of the samples added by *UH* were actually errors even though the classifier was confident they were correct.

The non-adapting classifier increased error over time for two of the TMR subjects and one able-bodied subject ($p<0.05$). Only two of the supervised adaptive classifiers increased error over time, and then only for a single subject ($p<0.05$). All of the other adaptive classifiers either reduced or maintained the same level of error over time.

DISCUSSION

On average, *Supervised High-confidence* adaptation *(SH)* added 62% of available samples, Supervised Low-confidence *(SL)* adaptation added 16% of available samples, and Supervised High/Low-confidence *(SHL)* adaptation added 75% of the available samples. *SL* provided equal reduction in error compared to *SH*, despite the fact that it only added 25% as many samples. It seems likely that, over time, repeatedly adding a large number of samples would lead to over-training, requiring a more selective inclusion criterion. *SL* likely preserves the responsiveness of the classifier to new adaptation by limiting the growth of the data set.

The Unsupervised High-confidence *(UH)* paradigm provided an implementable reduction in classification error. Although the <u>*Supervised*</u> Low-confidence paradigm provided a substantial increase in accuracy, the <u>*Unsupervised*</u> Low Confidence *(UL)* paradigm was unable to consistently predict the correct class. Future algorithms that are able to correctly identify the class in an unsupervised environment may be useful for long-term adaptation, but present implementations of *UL* are unacceptable for clinical implementation.

Some form of supervised adaptation should be incorporated into future clinically-viable algorithms. Although supervised adaptation paradigms require conscious training by the user, they are useful when the performance degrades to the point where the user is willing to push a button to go through a short training session to tune the classifier. The authors suggest that *SL* is the best supervised adaptation method, since it provides substantial reduction in error without adding a large number of samples.

Degradation Over Time. It is interesting to note that the only able-bodied subject whose non-adapting classification error increased over time was also the only able-bodied subject with previous experience controlling a pattern recognition system. All three of the TMR subjects also had extensive experience with a pattern recognition system, and two of them incurred increased error over time using the non-adapting classifier. There was no difference between experienced and inexperienced users regarding average non-adapting error, average minimum error, or average paired error, so it is unclear why experienced subjects seemed to incur more increase in error over time than inexperienced subjects. In a study that investigated the *UH* paradigm by Fukuda et al. [5], the non-adapting classifier significantly decayed over time, whereas there was no time degradation in the present study for many of the able-bodied subjects. Both studies used a single session that extended 1.5 – 2 hours. Fukuda et al. only used a remapped version of signal amplitude, which may be sensitive to fatigue. Classification of an ensemble of time-domain features has been shown to be robust to fatigue [3]. As a result, the ensemble of features used in this study may have prevented a need for fatigue-based adaptation due to the more robust set of selected features.

A linear discriminate analysis (LDA) classifier was chosen for this study, rather than a neural network as used in other adaptation studies [5]. Previous studies have shown that LDAs perform just as well as neural networks for static systems [7]. Although it is conceivable that a nonlinear classifier may have adapted better than an LDA, low levels of minimum error for each subject suggest that the LDA provided sufficient room for adaptation.

This study only involved a single two-hour session per subject. Studies that investigate adaptation over days or months require take-home prostheses capable of myoelectric pattern recognition. Once these prostheses are available, more practical results will be obtainable. The ability to design and evaluate adaptation algorithms is limited in part by our field’s conceptual weakness regarding what constitutes robustness or optimal performance. Adaptation work would greatly benefit from future research that provides a better mathematical and therapeutic framework through which to understand these critical concepts.

CONCLUSION

All supervised adaptation paradigms provided reduced classification error of a myoelectric system. Incorrect classification prevented unsupervised paradigms from achieving significant results, with the exception of a high-confidence unsupervised paradigm.

Acknowledgements: The authors thank Aimee E. Schultz, Kathy Stubblefield, and Laura Miller for reviewing this paper. This work was supported by the NIH National Institute of Child and Human Development (grant no. R01 HD043137).

REFERENCES

[1] D. Graupe and W. K. Cline, "Functional separation of EMG signals via ARMA identification methods for prosthesis control purposes," *IEEE transactions on Systems Man and Cybernetics,* vol. 2, p. 252:258, 1975.

[2] B. Hudgins, P. Parker, and R. N. Scott, "A New Strategy for Multifunction Myoelectric Control," *IEEE Transactions on Biomedical Engineering,* vol. 40, pp. 82-94, JAN 1993.

[3] M. T. Legere, J. R. Sexsmith, A. M. Mackie, R. N. Scott, and P. A. Parker, "Isometric muscle fatigue effects on pattern recognition ability of a new myoelectric prosthesis control system," in *Myoelectric Controls Symposium,* Fredericton, New Brunswick, 1993, pp. 10-16.

[4] L. Hargrove, K. B. Englehart, and B. Hudgins, "A training strategy to reduce classification degradation due to electrode displacements in pattern recognition based myoelectric control," *Biomedical Signal Processing and Control,* vol. doi:10.1016/j.bspc.2007.11.005, 2008.

[5] O. Fukuda, T. Tsuji, M. Kaneko, and A. Otsuka, "A human-assisting manipulator teleoperated by EMG signals and arm motions," *IEEE Transactions on Robotics and Automation,* vol. 19, pp. 210-222, APR 2003.

[6] D. Nishikawa, W. Yu, M. Maruishi, I. Watanabe, H. Yokoi, Y. Mano, and Y. Kakazu, "Online learning based electromyogram to forearm motion classifier with motor skill evaluation.," *JSME International Journal Series C-Mechanical Systems Machine Elements and Manufacturing,* vol. 43, pp. 906-915, 2000.

[7] L. Hargrove, K. Englehart, and B. Hudgins, "A Comparison of Surface and Intramuscular Myoelectric Signal Classification," *IEEE Transactions on Biomedical Engineering,* vol. 54, pp. 847-853, May 2007.

[8] A. B. Ajiboye and R. F. *ff.* Weir, "A heuristic fuzzy logic approach to EMG pattern recognition for multifunctional prosthesis control," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 13, pp. 280-291, SEP 2005.

[9] H. Huang, P. Zhou, G. L. Li, and T. A. Kuiken, "An analysis of EMG electrode configurations for targeted muscle reinnervation based neural machine interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 16, pp. 37-45, 2008.

[10] T. A. Kuiken, L. A. Miller, R. D. Lipschutz, B. A. Lock, K. Stubblefield, P. D. Marasco, P. Zhou, and G. A. Dumanian, "Targeted reinnervation for enhanced prosthetic arm function in a woman with a proximal amputation: a case study," *Lancet,* vol. 369, pp. 371-380, FEB 3 2007.

[11] T. A. Kuiken, G. A. Dumanian, R. D. Lipschutz, L. A. Miller, and K. A. Stubblefield, "The use of targeted muscle reinnervation for improved myoelectric prosthesis control in a bilateral shoulder disarticulation amputee.," *Prosthetics and Orthotics International,* vol. 28, pp. 245-253, December 2004.

[12] P. Zhou, M. Lowery, K. Englehart, L. Hargrove, J. P. A. Dewald, and T. A. Kuiken, "Decoding a New Neural-Machine Interface for Superior Artificial Limb Control," *Journal of Neurophysiology,* vol. 98, pp. 2974-2982, 2007.

[13] R. Kato, T. Fujita, H. Yokoi, and T. Arai, "Adaptable EMG Prosthetic Hand using On-line Learning Method -Investigation of Mutual Adaptation between Human and Adaptable Machine," in *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on,* 2006, pp. 599-604.