# Covariate Shift Adaptation in EMG Pattern Recognition for Prosthetic Device Control

Marina M.-C. Vidovic[1,2], Liliana P. Paredes[3], Han-Jeong Hwang[2], Sebastian Amsüss[5], Jaspar Pahl[1,4], Janne M. Hahne[2], Bernhard Graimann[6], Dario Farina[5], and Klaus-Robert Müller[2]

***Abstract*— Ensuring robustness of myocontrol algorithms for prosthetic devices is an important challenge. Robustness needs to be maintained under nonstationarities, e.g. due to electrode shifts after donning and doffing, sweating, additional weight or varying arm positions. Such nonstationary behavior changes the signal distributions – a scenario often referred to as covariate shift. This circumstance causes a significant decrease in classification accuracy in daily life applications. Re-training is possible but it is time consuming since it requires a large number of trials. In this paper, we propose to adapt the EMG classifier by a small calibration set only, which is able to capture the relevant aspects of the nonstationarities, but requires re-training data of only very short duration. We tested this strategy on signals acquired across 5 days in able-bodied individuals. The results showed that an estimator that shrinks the training model parameters towards the calibration set parameters significantly increased the classifier performance across different testing days. Even when using only one trial per class as re-training data for each day, the classification accuracy remained $>92\%$ over five days. These results indicate that the proposed methodology can be a practical means for improving robustness in pattern recognition methods for myocontrol.**

## I. INTRODUCTION

Pattern recognition methods have been extensively investigated as a means for controlling upper limb prostheses with EMG signals for the last decades [1]. Classifiers such as linear discriminant analysis (LDA) or support vector machine (SVM) [1], [2] show a high performance in myoelectric prosthesis control. However, these experiments were usually conducted under laboratory conditions. Unfortunately these controlled experimental conditions are of a limited relevance when translated into real prosthesis use in daily life activities. Laboratory tests indeed often do not include nonstationarities [3] and thus exhibit little robustness.

Factors that influence robustness include electrode shifts following donning and doffing, changes in arm position, variable loads when grasping objects, muscle fatigue, and varying electrode-skin impedances. An illustration of the potential impact of such nonstationarities in the example of electrode shifts is given in Fig. 1. A first attempt to mitigate the effect of those factors was investigated in [4]. In this and following studies, the main strategy has been the inclusion of examples of nonstationarities in the training set, thus increasing the generalization ability of the classifier to non-ideal conditions. Although this approach may improve the robustness, it requires massive (and often unfeasible) data recording efforts. Moreover, there are also factors, e.g. changes in user behavior, that cannot be easily included during training. Alternatively, it has been proposed to include a full calibration every day [1],which, however, is very demanding for the user.

We address the problem of robustness over changes in subject and recording conditions by adapting the algorithm to the daily conditions. This is accomplished by a short calibration set of signals that requires $< 1$ min for recording, so that the approach is practically feasible. We tested this approach on a large set of signals and the results showed that the proposed methodology is an appropriate compromise between a limited training time and a substantial improvement of classification accuracy.

## II. METHODS

### *A. Dataset*

The study was approved by the local ethics committees and involved six able-bodied subjects (five male, one female, age 25±2 yrs). For comparison, the subject group contained 2 experienced and 4 naive subjects. The data were recorded with eight commercially available double differential electrodes (13E200=50AC Otto Bock Healthcare Products GmbH, Vienna, Austria) placed equidistantly around the dominant forearm, approximately 7 cm to the olecranon.

*1) Experimental setup:* On five subsequent days, the subjects performed 8 movements: wrist pronation (WP), wrist supination (WS), wrist extension (WE), wrist flexion (WF), hand opening (HO), fine pinch (FP), key grip (KG), and no movement (NM). On the basis of

*This work has been supported by the German Ministry for Education and Research (BMBF) via the Bernstein Focus Neurotechnology (BFNT) Göttingen and the European Commission via the Industrial Academia Partnerships and Pathways (project AMYO, Grant No. 251555.)

[1]Otto Bock Healthcare GmbH, Kaiserstr. 39, 1070 Vienna, Austria marina.vidovic@ottobock.com

[2] Machine Learning Group, Technical University of Berlin (TU Berlin), Marchstr. 23, 10587 Berlin, Germany

[3] Laboratorio di Robotica e Cinematica, Via Alberoni 70, 30126 Venezia, Lido, Italy

[4] Faculty of Electrical Engineering and Information Technology, Otto-von-Guericke University of Magdeburg, 39106 Magdeburg, Germany

[5] Department of Neurorehabilitation Engineering, Georg August University, 37073 Göttingen, Germany

[6]Otto Bock Healthcare GmbH, 37115 Duderstadt, Germany

978-1-4244-7929-0/14/$26.00 ©2014 IEEE 

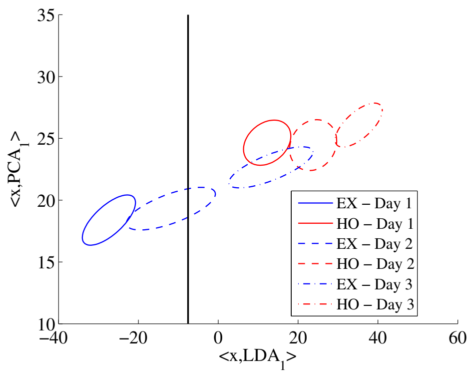



Fig. 1. Illustration of the possible effect of electrode shifts for the classification between two movements: Hand Open (HO) and Extension (EX). The optimal separation hyperplane for the first day data, shown as black vertical line, performs worse on the data of the second day and fails completely on the data of the third day. For a two dimensional representation, the data were projected on the first LDA and PCA component, respectively [5].

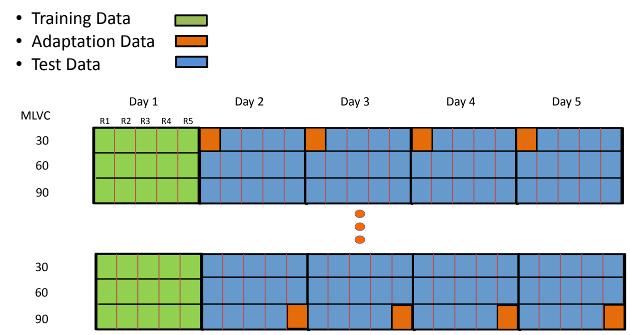



Fig. 2. Illustration of the data segmentation, where $R_n$, $n = 1,\ldots,5$ includes the trials of all movements for one particular contraction level. The classifier was trained on one day (exemplary on Day 1 (green)), adapted by one run of the test day (orange) and tested on all other runs of the same day (blue).

visual feedback, the subject performed each movement at three contraction forces (30, 60, and 90% maximum long-term voluntary contraction, MLVC). We recorded five trials for each movement and each contraction level. This led to a total of 8 classes $\times$ 3 contraction levels $\times$ 5 repetitions = 120 trials for each day. One trial comprised 5 s of data recording. In the first second, the subject increased the muscle force to the target force, then he/she maintained the target force for 3 s, followed by relaxation. For the following analyses, we used only the 3-s recording intervals where the subject maintained a constant force level. On the first day, the electrode positions were marked with a pen on the skin, so that the same positions could be reproduced across days. Fig. 2 illustrates the data segmentation, where $R_n$, $n = 1,\ldots,5$ includes the trials of all movements for one particular contraction level.

*2) Signal acquisition and processing:* The acquired raw signals were amplified to the range 0-4.5 V and filtered in the bandwidth 20-450 Hz, with the inclusion of a 50-Hz notch filter by the active Otto Bock electrodes. The data were sampled at 1 kHz, digitized by a 10 bit A/D converter, and transferred to a computer via Bluetooth by the Axonmaster (Otto Bock HealthCare Products GmbH, Vienna, Austria).

*3) Feature extraction:* As proposed in [6], the logarithm of the signal variance (logVAR) was calculated from the raw EMG data. The logVAR features were computed from 250-ms intervals, which overlapped by 50 ms (15 samples per trial).

### *B. Mathematical Background*

In this study, we focused on multi-class classification problems. We consider two Bayesian Classifiers: quadratic discriminant analysis (QDA), that determines quadratic decision boundaries and linear discriminant analysis (LDA), that uses hyperplanes for class separation. For an optimal classification we would classify an unknown point $x$ according to the largest posterior probability [7]. Since in practice the posterior probabilities cannot be obtained directly, we estimate them using the training data. Let $X = \{(x_i, y_i)_{i=1}^n\}$, $n \in \mathbb{N}$ be the training samples, where $x_i \in \mathbb{R}^d$, $d \in \mathbb{N}$ and $y_i \in \{1,\ldots,C\}$. Given the prior probability $\pi_c = 1/C$ for each class $c \in \{1,\ldots,C\}$ and the class-conditional density functions $f_c(x)$ of $X$, we can estimate the class posteriors probabilities by Bayes Rule

$$Pr(c|x) := \frac{f_c(x)\pi_c}{\sum_{l=1}^{C} f_l(x)\pi_l}. \tag{1}$$

Now, we assume that the data are Gaussian distributed and substitute $f_c(x)$ with the multivariate Gaussian distribution

$$p(x,c) = \frac{1}{\sqrt{(2\pi)^d|\Sigma_c|}} \exp\big(-\frac{1}{2}\hat{x}^\top\Sigma_c^{-1}\hat{x}\big), \tag{2}$$

where $\hat{x} := x - \mu_c$. By taking additionally the natural logarithm of Equation (1), the quadratic discriminative function (QDA) $\delta_c^1$ for each class $c$ results in

$$\delta_c^1(x) := -\frac{1}{2}\log|\Sigma_c| - \frac{1}{2}\hat{x}^\top\Sigma_c^{-1}\hat{x} + log\pi_c. \tag{3}$$

Assuming further that all classes share the same covariance matrix, Eq. (3) can be rearranged to the linear discriminant function (LDA) $\delta_c^2$

$$\delta_c^2(x) := x^\top\Sigma^{-1}\mu_c - \frac{1}{2}\mu_c^\top\Sigma^{-1}\mu_c + log\pi_c, \tag{4}$$

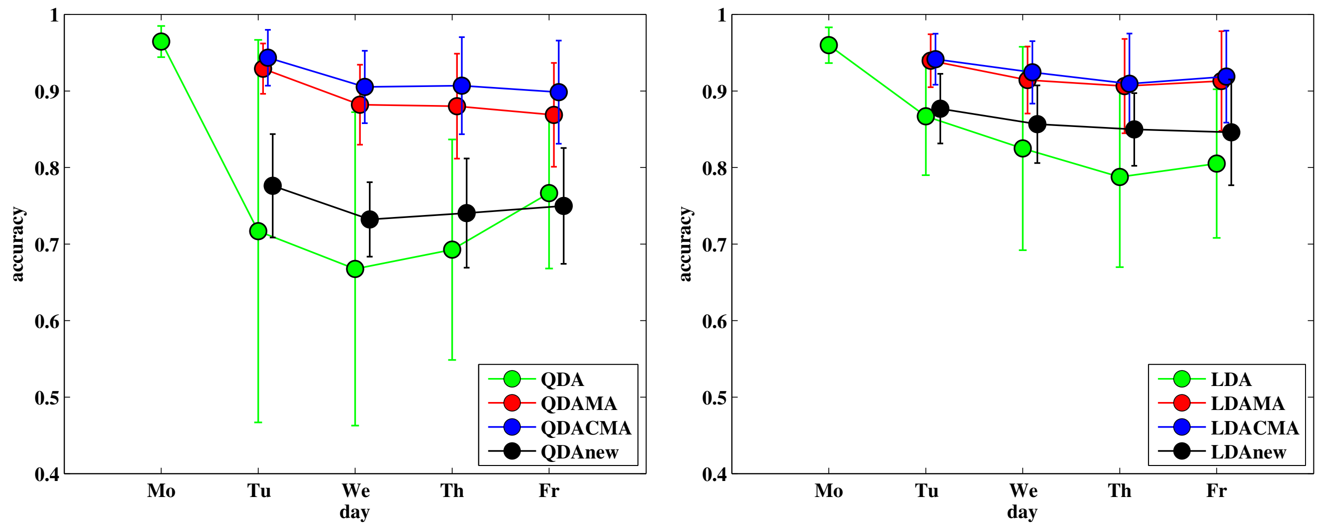
where $\Sigma = \frac{1}{C}\sum_{c=1}^{C}\Sigma_c$ is the pooled covariance matrix.

Finally, for applying QDA ($j = 1$) and LDA ($j = 2$), an unknown point $x^* \in \mathbb{R}^d$ is allocated to the class with the highest probability, respectively

$$y^* = \arg\max_c \delta_c^j(z), \tag{5}$$

where $y^* \in \{1,\ldots,C\}$.

However, due to the previously discussed nonstationarities and the resulting covariate shift (cf. [8]), the test distribution $p_{te}(x)$ potentially differs from the training distribution $p_{tr}(x)$: $p_{tr}(x) \neq p_{te}(x)$. In this situation, the




(a) QDA

(b) LDA

Fig. 3. Illustration of a) QDA and b) LDA adaptation performance. In both cases, applying the mean adaptation results in a remarkable improvement of classification accuracy across days. Additional inclusion of the adapted covariance matrices only results in a slight improvement in the case of QDA. Solely using the new calibration data for training of a completely new classifier slightly improves the performance but performs worse than the adapted classifiers.

trained decision boundaries potentially are no longer ensuring a satisfactory class separation and an adjustment of the trained classifier to the changed data distribution is needed. Accordingly, in the following we propose an adaptation methodology, that, given a small calibration data set, adjusts the training model parameters towards the calibration set parameters.

### C. Adaptation

The adaptation was done by a small calibration data set $X_{cal} = \{t_i, u_i\}_{i=1}^m,\ m \in \mathbb{N}$, which follows the test distribution $X_{cal} \sim p_{te}$, where $t_i \in \mathbb{R}^d$ and $u_i \in \{1, \dots, C\}$.

Let $\mu_{cal_c}$ and $\Sigma_{cal_c}$ be the mean and the covariance matrix of $X_{cal}$ and $\mu_{tr_c}$ and $\Sigma_{tr_c}$ the ones of the training set $X$. We introduce an adaptation by shrinking the training parameters towards the ones obtained from the calibration set:

$$\tilde{\mu}_c = (1 - \tau)\mu_{tr_c} + \tau\mu_{cal_c}, \tag{6}$$

and

$$\tilde{\Sigma}_c = (1 - \lambda)\Sigma_{tr_c} + \lambda\Sigma_{cal_c}, \tag{7}$$

where $\tau$ and $\lambda \in [0, 1]$ are the regularization parameters. The optimal values for $\tau$ and $\lambda$ where estimated by a grid search with a step sizes of 0.1 across all days and subjects. In case of optimizing only one shrinkage parameter, the other is held at zero.

### D. Evaluation procedure

The proposed adaptation scheme was evaluated offline with a reverse leave-one-out cross validation. The data were split as shown in Fig. 2 and the entire data set of one day (green) was used as initial classifier training set. For testing on different days, one trial for each class (orange), denoted as run, was used as calibration set $X_{cal}$ (thus, a total of 24 s of data including all classes). All other trials were used as test set (blue) for which the classification accuracy was calculated. In this way, we proceeded for each run and computed the average over all results.

## III. RESULTS

The results are presented as average over the six subjects.

### A. Covariate shift adaptation

A QDA classifier was trained on the complete data set of the first day and tested on all the other days (green in Fig. 2). Starting with an accuracy $>$ 95% on the same day as training, there was a decrease in accuracy of almost 30% on the next days. The QDA performance dropped substantially when tested on the 2nd to 4th day and recovered only slightly on the last day. The results are shown as a green curve in Fig. 3 a). The optimal values for $\lambda$ and $\tau$ were determined by a grid search and the results are shown in Fig. 4.

Firstly, a QDA adaptation by shrinking only the mean of the QDA to the mean of the calibration set was evaluated (QDAMA: ($\tau = 0.8,\ \lambda = 0$)). In a second investigation the covariance matrix was additionally adapted and we denote this adaptation as QDA covariance mean adaptation (QDACMA: ($\tau = 0.8,\ \lambda = 0.7$)). Lastly, QDAnew ($\tau = 1,\ \lambda = 1$) denotes the classifier being trained only on the small calibration data set to compare the adaptation performance. With respect to the QDA performance the mean adaptation (QDAMA), displayed as red curve in Fig. 3 a), considerably improved the classification accuracy for each day. Furthermore, an additional performance gain was obtained by adapting the covariance matrix (QDACMA), even though the increase with respect to QDAMA was relatively small.

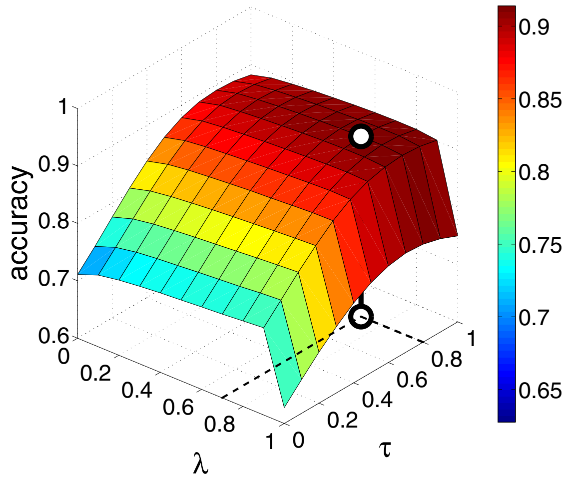



Fig. 4. Illustration of the grid search for finding the optimal adaptation parameters $\tau$ and $\lambda$ for QDA. A continuous performance increasing was preserved by shrinking the mean of the training set versus the mean of the calibration set, where the optimal value was $\tau = 0.8$. Shrinking additionally the covariance matrix by $\lambda = 0.7$ a further, but relatively small performance gain was obtained.

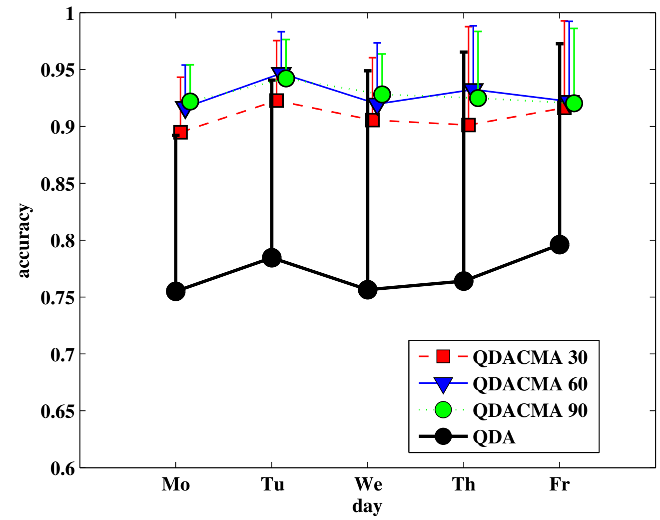



Fig. 5. A comparison of the QDACMA performance when using a calibration set with 30, 60 and 90% contraction strength. Adapting the QDA by a calibration set with 60 or 90% contraction strength indicated the best performance.

QDAnew was on average better than the old QDA but worse than both QDA adaptation variations.

The same adaptation experiments were repeated for LDA, where the optimal parameters for the mean adaptation (LDAMA: $\tau = 0.7$, $\lambda = 0$) and for the additional covariance matrix adaptation (LDACMA: $\tau = 0.6$, $\lambda = 0.7$) were determined by using a grid search as shown in Figure 4. The results are shown in Fig. 3 b). The behaviors of the adaptation methods were similar to the ones of QDA except that adapting the covariance matrix additionally to the mean did not change the performance notably.

### *B. Influence of force contraction level*

It was further investigated, which contraction level, i.e. 30, 60 or 90% MLVC, should be used for adaptation. Therefore the above experiment (shown in Fig. 3) was repeated for each contraction level separately. Additionally, each day was used once for training and all other days for testing. The mean performance for the test days is shown in Fig. 5. All adaptation scenarios outperformed the non-adapted QDA. Furthermore, an adaptation with 60 or 90% MLVC was the best choice.

## IV. DISCUSSION AND CONCLUSIONS

Among the many challenges in myocontrol of prostheses, we increased robustness over repeated uses by an adaptive method, that alleviates the effects of nonstationarities. For this purpose, we proposed an adaptation scheme that only requires a 1-min recording of new labeled data each day. The results demonstrated the gain in classification accuracy when using the proposed methodology in comparison to the non-adapted classifier in an offline analysis for able-bodied subjects. Due to the extensive training (2 hours for 5 days) in this study, on average no absolute recognition accuracy difference was found between the experienced and inexperienced subject group. Further, the relative improvement trends achieved with the proposed adaptive approach were the same for all subjects, underlining the relevance of our method for both experienced and novel users. Further studies on users with transradial amputation are currently being performed with promising preliminary results.

## ACKNOWLEDGMENT

We greatly thank Nico Görnitz and Daniel Bartz for interesting discussions and helpful comments.

## 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.," *Journal of Rehabilitation Research & Development*, vol. 48, no. 6, 2011.

[2] M.-F. Lucas, A. Gaufriau, S. Pascual, C. Doncarli, and D. Farina, "Multi-channel surface emg classification using support vector machines and signal-based wavelet optimization," *Biomedical Signal Processing and Control*, vol. 3, no. 2, pp. 169–174, 2008.

[3] N. Jiang, S. Dosen, K.-R. Müller, and D. Farina, "Myoelectric control of artificial limbs-is there a need to change focus," *IEEE Signal Processing Mag.*, vol. 29, no. 5, pp. 149–152, 2012.

[4] J. W. Sensinger, B. A. Lock, and T. A. Kuiken, "Adaptive pattern recognition of myoelectric signals: exploration of conceptual framework and practical algorithms," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 17, no. 3, pp. 270–278, 2009.

[5] P. Shenoy, M. Krauledat, B. Blankertz, R. P. Rao, and K.-R. Müller, "Towards adaptive classification for BCI," *Journal of neural engineering*, vol. 3, no. 1, p. R13, 2006.

[6] J. Hahne, F. Biessmann, N. Jiang, H. Rehbaum, D. Farina, F. Meinecke, K.-R. Müller, and L. Parra, "Linear and nonlinear regression techniques for simultaneous and proportional myoelectric control," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 22, no. 2, pp. 269–279, 2014.

[7] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning*, vol. 2. Springer, 2009.

[8] M. Sugiyama, M. Krauledat, and K.-R. Müller, "Covariate shift adaptation by importance weighted cross validation," *The Journal of Machine Learning Research*, vol. 8, pp. 985–1005, 2007.