A Benchmark Database for Myoelectric Movement Classification

Journal:	*Transactions on Neural Systems & Rehabilitation Engineering*
Manuscript ID:	TNSRE-2013-00051
Manuscript Type:	Paper
Date Submitted by the Author:	14-Mar-2013

Complete List of Authors:	Atzori, Manfredo; University of Applied Sciences Western Switzerland, Sierre (HES-SO Valais), Gijsberts, Arjan; Institute de Recherche Idiap, Kuzborskij, Ilja; Institute de Recherche Idiap, Heynen, Simone; University of Applied Sciences Western Switzerland (HES-SO Valais), Physical Therapy Mittaz Hager, Anne-Gabrielle; University of Applied Sciences Western Switzerland (HES-SO Valais), Physical Therapy Deriaz, Olivier; Clinique romande de réadaptation Suvacare, Service de recherche et contrôle qualité médicale Castellini, Claudio; DLR, RM Müller, Henning; University of Applied Sciences Western Switzerland, Sierre (HES SO), Business Information Systems Caputo, Barbara; Institute de Recherche Idiap,

TIPS:	surface electromyography, hand prosthetics, hand movements classification, public databases

A Benchmark Database for Myoelectric Movement Classification

Manfredo Atzori, Arjan Gijsberts, Ilja Kuzborskij, Simone Heynen, Anne-Gabrielle Mittaz Hager, Olivier Deriaz, Claudio Castellini, Henning Müller, and Barbara Caputo

***Abstract*—We describe the NINAPRO database, a publicly available resource that aims to support research on advanced myoelectric hand prosthetics. The database is obtained by jointly recording surface electromyography signals from the forearm and kinematics of the hand and wrist while subjects perform a predefined set of actions and postures. The current milestone release of the database contains data obtained from 27 intact subjects performing 52 finger, hand, and wrist movements. Additional data acquisitions from both intact and amputated subjects are ongoing and will be added periodically to the database. Besides describing the acquisition protocol and processing procedures, we also present benchmark classification results using a variety of feature representations and classifiers. Statistical analysis on the these results provides empirical evidence that classification accuracy is negatively correlated with the subject's Body Mass Index.**

***Index Terms*—**

I. INTRODUCTION

Since the 1960s, pattern recognition algorithms have been applied on surface electromyography (sEMG) signals to control simple mechanical grippers with a single Degree of Freedom (DOF) [1, 2, 3]. The principal goal of this research was to predict the intent of an amputee and to use this to control a dexterous, self-powered hand prosthesis. Indeed, an amputee should be able to dexterously control the prosthesis just by desiring to so in a natural way. Still, 45 years later this goal has not yet been reached: one quarter to one third of the amputees reject self-powered prostheses due to low reliability, weight, trouble with maintenance, low dexterity, and poor visual appearance [4, 5].

A major obstacle towards this goal is the lack of a standard benchmark for sEMG-based control of hand prostheses. To the best of our knowledge, all studies in this field have been performed using proprietary data and are limited to groups that possess the equipment, expertise, and manpower to acquire the necessary data. As a consequence, the specific application domain is not widely accessible for researchers of other fields, such as machine learning or signal processing. Moreover, the scale of acquisition is often limited to the minimum required to verify a specific scientific hypothesis— this usually means a dozen intact subjects or a few amputees. Last but not least, there is no standard for experimental setups and protocols (e.g., the set of movements, electrode placement), nor are standardized databases publicly available. This is in contrast to several other research communities, where wide acceptance of common, publicly available benchmark databases has considerably pushed progress and helped to identify open challenges. This has been the case for the fields of computer vision (e.g., PASCAL [6], CALTECH 256 [7], SUN [8]), robotics (e.g., Radish [9], RGB-D SLAM [10]), medical informatics (e.g., ImageCLEF [11]), as well as many others. This lack of standard benchmarks decreases the reliability of research results and reduces the possibility that new techniques can be applied successfully in commercial applications. The diversity in experimental setups and protocols makes it infeasible to compare results among different studies, making it hard to evaluate whether certain approaches are actually to be preferred over others.

We believe that the time is ripe for the biorobotics community to have such a benchmark. To this end, we present the milestone release of the Non-Invasive Adaptive Prosthetics (NINAPRO) database, which was succinctly introduced by Atzori et al. [12]. The database is presented jointly with the acquisition setup and protocol, data processing routines, and characteristics of the subjects involved in the data acquisition. The actual data consists of sEMG and kinematic signals of the wrist and hand gathered from 27 intact subjects performing 52 hand movements. These movements were selected from the relevant literature and standard rehabilitation practice guidelines. Collection of the database is a continuous and ongoing effort, and additional acquisitions from intact as well as amputated subjects will be added.

Aside from describing the acquisition procedure, we also present an extensive benchmark comparison using a large variety of popular feature extraction and classification methods. The results indicate that non-linear classifiers are required to successfully discriminate all 52 movements. Furthermore, regression analysis on the classification results reveals that classification accuracy is negatively correlated with a subject's Body Mass Index (BMI).

A detailed description of the acquisition setup and protocol follows in Section II. Data processing routines will subsequently be presented in Section III, which includes

M. Atzori and A. Gijsberts contributed equally to this work.

M. Atzori and H. Müller are with the Department of Business Information Systems at the University of Applied Sciences Western Switzerland (HES-SO Valais), Sierre, Switzerland.

A. Gijsberts, I. Kuzborskij, and B. Caputo are with the Institute de Recherche Idiap, Martigny, Switzerland.

S. Heynen and A.-G. Mittaz Hager are with the Department of Physical Therapy at the University of Applied Sciences Western Switzerland (HES-SO Valais), Leukerbad, Switzerland.

O. Deriaz is with the Institut de recherche en réadaptation, Service de recherche et contrôle qualité médicale, Clinique romande de réadaptation, Suvacare, Sion, Switzerland.

C. Castellini is with the Institute of Robotics and Mechatronics of the DLR - German Aerospace Research Center, Wessling, Germany.

a principled relabeling strategy to correct erroneous labels. Section IV presents a benchmark evaluation on the database as well as a multiple regression analysis to investigate which subject properties affect classification accuracy. Conclusions and future work are subsequently covered in Section V.

II. Building the Ninapro Database

A. Acquisition setup

The Ninapro database in its present release combines kinematic hand and wrist data, acquired using a CyberGlove and an inclinometer, with muscular activity data acquired using Otto Bock sEMG electrodes. All devices are certified according to medical and electric safety standards in the United States and the European Union.

1) Surface electromyography: The muscular activity is gathered using ten active double-differential OttoBock *MyoBock* 13E200 sEMG electrodes[1], which provide an amplified, bandpass-filtered, and Root Mean Square (RMS) rectified version of the raw sEMG signal. The electrodes' amplification is set to a factor of 14000. Particular care was taken when deciding the *placement* of the electrodes on the forearm or stump. Choosing the right positions of the electrodes is usually regarded as a crucial step and several attempts have been made at targeting forearm muscles on healthy subjects [2, 13] as well as on amputees [14]. However, early research on pattern recognition for sEMG [15, 16] (recently confirmed in [17]) proved that targeted placement of electrodes is not required when doing posture classification, since pattern recognition techniques can compensate for suboptimal placement and may even take advantage of muscle cross-talk.

Eight electrodes are uniformly placed around the forearm or stump using an elastic band, at a constant distance from the radio-humeral joint just below the elbow. Two additional electrodes are placed on the large flexor and extensor muscles of the forearm (see Figure 1). This positioning of the electrodes also gives the opportunity to improve classification results by applying linear and non linear spatial registration algorithms, as described in Atzori et al. [18].

2) Kinematics: The kinematic configuration of the hand is measured using a 22-sensor CyberGlove II dataglove[2], shown in the right panel of Figure 1. The CyberGlove is a light fabric, elastic glove, onto which 22 strain gauges are sewn. The sewing sheaths are chosen carefully by the manufacturer, so that the gauges exhibit a resistance that is proportionally related to the angles between pairs of hand joints of interest. The device returns 22 8-bit values proportional to these angles for an average resolution of less than one degree depending on the size of the subject's hand, careful wearing of the glove, and the angular range of the considered joint. Note that we record raw sensor values rather than estimated joint angles, the reason being that reliable calibration of the glove is prohibitively time-consuming. In addition, most machine learning techniques are invariant to linear scaling of the data and calibration is in these cases unnecessary. If desired, exact joint angles can be obtained by calibrating the glove a posteriori for a given subject.

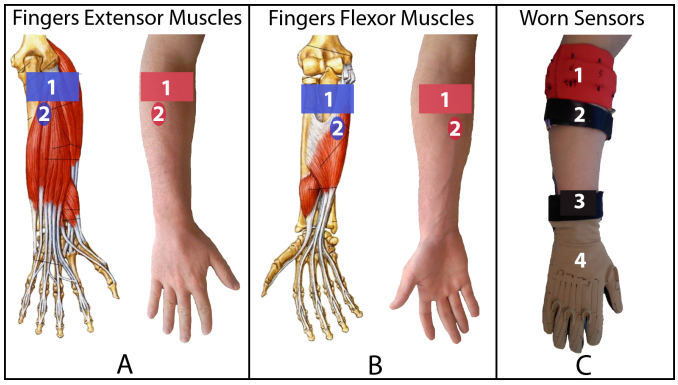


Fig. 1. Placement of the electrodes: A. sEMG electrodes placed on finger extensor muscles (A.1 Equally spaced electrodes; A.2 Spare electrode); B. sEMG electrodes placed on finger flexor muscles (B.1 Equally spaced electrodes; B.2 Spare electrode); C. all the sensors positioned on the arm (C.1 Equally spaced electrodes; C.2 Spare electrode; C.3 Inclinometer; C.4 CyberGlove II);

In addition to the CyberGlove, a standard commercially available 2-axis Kübler IS40 inclinometer[3] is fixed onto the subject's wrist to measure the wrist orientation. The inclinometer has a range of 120° and a resolution of less than 0.15°.

3) Data acquisition: Data from the electrodes and the inclinometer are acquired at a constant interval of 100 Hz using a standard National Instruments DAQ card (NI-DAQ PCMCIA 6024E, 12-bit resolution). Kinematic data from the CyberGlove are recorded over a Bluetooth-tunneled serial port at slightly less than 25 Hz. Each data sample is associated with an accurate timestamp (using Windows Performance Counters) and directly written to mass storage.

B. Experimental protocol

Preceding the experiment, each subject is requested to give informed consent and to fill in a brief questionnaire concerning clinical data. These data include age, gender, height, weight, fitness, laterality, and self-reported health status. In the case of amputees, we also note the age, type, and reason of the amputation; information about the use of prostheses (cosmetic, body-powered, self-powered, etc.) along with the (dis)advantages and consequences of their usage; type and degree of phantom limb sensation and pain. Moreover, we take pictures of the arm with and without the acquisition setup to be able to check macroscopic differences. After finalizing the forms, the subject is asked to sit comfortably on an adjustable chair in front of a table with a large monitor, while the sEMG electrodes, dataglove, and inclinometer are worn on the right arm. Amputees wear the sEMG electrodes on the stump instead, while the dataglove and the inclinometer are worn on the intact limb. The subjects are asked to bilaterally perform the movement shown on the screen according to the bilateral imitation procedure [17].

First, the subjects have to perform a training sequence that involves three repetitions of a selection of movements

[1]Otto Bock HealthCare GmbH, http://www.ottobock.com
[2]CyberGlove Systems LLC, http://www.cyberglovesystems.com
[3]Fritz Kübler GmbH, http://www.kuebler.com

in order to get accustomed to the protocol. Then, the real data acquisition starts and the subjects have to repeat ten times the 52 movements. Each movement repetition lasts 5 s and is followed by 3 s of rest. The movements have been selected from the hand taxonomy and robotics literature (see, e.g., [19, 20, 21, 22, 23]), as well as from the *Disabilities of the Arm, Shoulder and Hand protocol* for functional movements [23].

The experiment is divided into three exercises:

1) 12 basic movements of the fingers (flexions and extensions);
2) 8 isometric and isotonic hand configurations and 9 basic movements of the wrist (adduction/abduction, flexion/extension, and pronation/supination); and
3) 23 grasping and functional movements—in this case, everyday objects are presented to the subject for grasping, in order to mimic a daily-life action.

See Table I and Figure 2.

The exercises last respectively 16, 23, and 31 minutes. Subjects are allowed short breaks between exercises to avoid muscle fatigue, such that the total duration of the experiment is around 100 minutes (including preparatory steps and training). The sequence of movements is not randomized in order to induce unconscious movement repetitions into the subjects. The experiment received the approval of Ethics Commission of the Canton of Valais (Switzerland), where all acquisitions have been performed.

TABLE I
SYNTHETIC DESCRIPTIONS OF THE 52 MOVEMENTS OF INTEREST, ALONG WITH A REFERENCE, IF AVAILABLE.

	#	Description	Ref.
Finger mvts.	1-2	Index flexion and extension	[14]
	3-4	Middle flexion and extension	[14]
	5-6	Ring flexion and extension	[14]
	7-8	Little finger flexion and extension	[14]
	9-10	Thumb adduction and abduction	[14]
	11-12	Thumb flexion and extension	
Hand postures	1	Thumb up	[24]
	2	Flexion of ring and little finger; thumb flexed over middle and little	
	3	Flexion of ring and little finger	[25]
	4	Thumb opposing base of little finger	[25]
	5	Abduction of the fingers	[25]
	6	Fingers flexed together	[25]
	7	Pointing index	[26]
	8	Fingers closed together	[27]
Wrist mvts.	1-2	Wrist supination and pronation (rotation axis through the middle finger)	[24]
	3-4	Wrist supination and pronation (rotation axis through the little finger)	
	5-6	Wrist flexion and extension	[24]
	7-8	Wrist radial and ulnar deviation	[27]
	9	Wrist extension with closed hand	
Grasping and functional movements	1-2	Large and small diameter	[19]
	3	Fixed hook	[19]
	4	Index finger extension	[19]
	5	Medium wrap	[19]
	6	Ring	[19]
	7	Prismatic four fingers	[19]
	8	Stick	[19]
	9	Writing tripod	[19]
	10-12	Power, three finger, and precision sphere	[19]
	13	Tripod	[19]
	14-15	Prismatic and tip pinch	[19]
	16	Quadpod	[19]
	17	Lateral	[19]
	18-19	Parallel extension and flexion	[19]
	20	Power disk	[19]
	21	Open a bottle with a tripod grasp	[23]
	22	Turn a screw (grasp the screwdriver with a stick grasp (8))	
	23	Cut something (grasp the knife with an index finger extension grasp (4))	[23]

C. *The* NINAPRO *database*

Finally, the data of each subject are stored anonymously in a database and made available on a website[4]. The first version of the database contains data of 27 intact subjects (20M/7F, 25/2 right-/left-handed, age 28.0 ± 3.4 y). For each subject and each exercise, three data files are stored in plain ASCII format, containing the signals from (1) the electrodes and the inclinometer, (2) the cyberglove, and (3) the video stimulus. The data are arranged in a line, consisting of a timestamp plus the sensor values. The clinical data and five pictures (three previews of the data of each exercise plus two pictures of the forearm and of the hand with and without the acquisition setup) are also stored for each subject.

III. DATA PROCESSING

The acquisition software described in the previous section stores individual modalities to separate files. Most practical applications, however, require further processing of the data. These processing steps include synchronization of the kinematic and sEMG streams with the stimulus, removal of noise components by filtering the data, and finally correcting mislabeled samples using a relabeling strategy. The described processing steps are the result of initial experiments and intended specifically to support the analyses in Section IV, while the publicly available NINAPRO database contains the raw data as described in the previous section. Implementations of our processing methods are available separately upon request.

[4]The database can be accessed (after the request of credentials) at http://ninapro.hevs.ch. Reviewers can access this resource using the username "Reviewer" and password "tnsre2013". Supporting files for the acquisition setup and protocol (e.g., software and stimulus videos) can be obtained on an individual basis by contacting the authors.

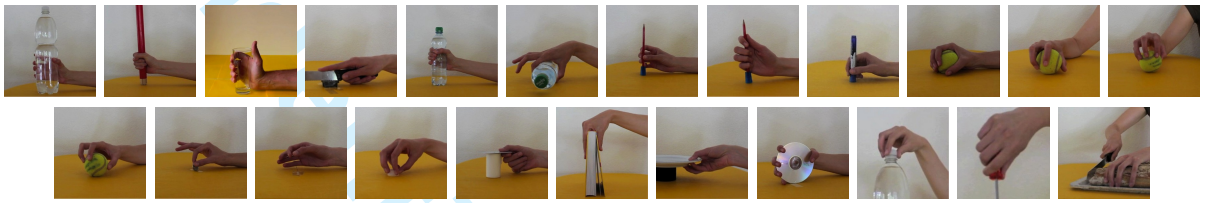
A. *Synchronization and Filtering*

Synchronization of the input modalities is relatively straightforward with the NINAPRO database, since each datum is recorded with an accurate timestamp. The difference in sampling rates is eliminated by linearly interpolating all the data streams to the highest recording frequency (i.e., 100 Hz for the sEMG stream). The following processing step is to low-pass filter the sEMG signals at a cutoff frequency of 5 Hz using a zero-phase second order Butterworth filter. This low cutoff frequency is justified in our setting, since the RMS filtering onboard the Otto Bock electrodes drastically changes the spectral properties of the signal. In contrast, for raw sEMG recordings the relevant spectral domain is typically reported as approximately between 15 to 500 Hz.

(a) Basic movements of the fingers (flexions and extensions).

(b) Isometric, isotonic hand configurations ("hand postures").

(c) Basic movements of the wrist.



(d) Grasping and functional movements.

Fig. 2. The 52 movements of interest.

B. Relabeling

Human reaction times and attention spans inevitably cause some misalignment between the stimulus video and the actual movement as performed by the subject. An example of this misalignment is shown in the left panel of Figure 3, which overlays the movement label imposed by the video (marked as "movement") on top of the sEMG activity. In this case, the subject finishes the movement well before the video ends. A considerable number of samples near the end are erroneously marked as posture, while the subject in fact already returned to the rest position. To reduce this label "noise", we devise an offline relabeling algorithm that constrains movement labels to those samples in which there is increased sEMG activity.

Similar to the onset detection approach by Staude [28], we first remove irrelevant autoregressive components by whitening the rectified signals using a multivariate VAR(p) model [29]. In our case, an order of $p = 20$ was found to perform adequately. Detection of sEMG activity is restricted to the original video window extended with an additional 100 samples at the end to allow subjects to finish a movement with up to 1 s of delay. The resulting feasible movement window (see Figure 3, center) of length T is then divided in rest-movement-rest segments marked by change points t_0 and t_1.

The optimal change points are found by maximizing the Generalized Likelihood Ratio (GLR) between the rest model $\boldsymbol{\theta}_0$ and movement model $\boldsymbol{\theta}_1$. The corresponding objective function can be written as

$$\underset{1\le t_0\le T}{\arg\max}\ \underset{t_0\le t_1\le T}{\arg\max}\ \sup_{\boldsymbol{\theta}_0\in\Theta_0}\ \sup_{\boldsymbol{\theta}_1\in\Theta_1} \left[\sum_{i=1}^{t_0-1}\ln p_{\boldsymbol{\theta}_0}(\boldsymbol{y}_i)+\sum_{j=t_0}^{t_1-1}\ln p_{\boldsymbol{\theta}_1}(\boldsymbol{y}_j)+\sum_{k=t_1}^{T}\ln p_{\boldsymbol{\theta}_0}(\boldsymbol{y}_k)\right]. \tag{1}$$

Simple exhaustive search is adequate for finding optimal t_0 and t_1, while $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ are optimized by a maximum likelihood estimate of a multivariate Gaussian distribution over the corresponding window segments.

To improve segmentation on noisy data, we also impose a minimum duration for both the rest (i.e., $t_0 \geq 10$) and movement window segments (i.e., $t_1 - t_0 \geq 0.3T$). Moreover, the reasonable assumption that sEMG activity is higher during movements than during rest is explicitly enforced by requiring the sample variance s^2 to be higher during movements (i.e., $s_1^2 \geq s_0^2$). This simple condition is effective at preventing erroneous outcomes in cases where a feasible window is lacking a clear initial rest. Finally, we impose a prior distribution on any sample belonging either to rest or movement (i.e., random variables R_i and M_i). This prior is chosen uniformly as $p(R_i) = 0.1$ for $1 \leq i \leq T$, and due to mutual exclusivity $p(M_i) = p(\neg R_i) = 1 - p(R_i)$. The effect of this prior is that the algorithm will identify slightly larger movement windows, which helps to ensure that the entire sEMG activity is captured in the movement segment.

IV. Analysis

Jointly with the acquisition protocol and the corresponding database, we also present a classification benchmark obtained with various feature extraction and classification methods.

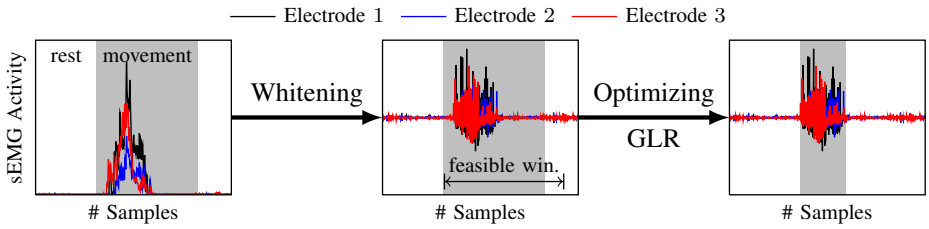


Fig. 3. Illustration of the relabeling process consisting of whitening the rectified sEMG signal and optimization of the Generalized Likelihood Ratio within the feasible window. The shaded area indicates time windows labeled as a non-rest movement. For clarity the figure displays only three out of ten electrodes.

These results are not only intended as a reference for future work, but also to provide suggestions on which methods perform particularly well with the given data. The comparatively large number of subjects in the database allow us to subsequently determine the influence of various subject properties on classification accuracy.

A. Classification Benchmark

Successful classification of movements from sEMG signals depends to large extent on the type of feature representation and classifier. To ensure that we establish (near) optimal performance, we consider a variety of popular feature extraction and classification methods. Aside from producing a direct quantitative comparison of these methods, this benchmark also investigates whether current state-of-the-art methods can indeed attain satisfactory levels of performance on this challenging setting.

1) Methods and Experimental Setup: Feature representations used on sEMG signals can roughly be divided in three categories, namely representations in the time domain, in the frequency domain, and finally representations that relate to both time and frequency domains [30, 5]. Among these, representations in the time domain have traditionally been popular for sEMG signals, due to ease of computation and since they reduce a processing window to a scalar value. As representatives of these simple time domain features we consider Mean Absolute Value (MAV)[5], the Variance (VAR), and the Waveform Length (WL). A potential shortcoming of these methods is that the drastic reduction to a scalar value leads to a loss of information. As alternative representations that preserve more information we therefore also include the sEMG Histogram (HIST) and Cepstral Coefficients (CC). The latter method applies a logarithm on the spectral coefficients and subsequently maps the data back into the time domain by means of an inverse Fourier Transform.

Frequency-domain features are commonly based on the Fourier Transform, of which we consider the Short-Time Fourier Transform (STFT) variant. An alternative representation that has recently gained popularity is the Discrete Wavelet Transform (DWT). This transformation decomposes the signal in terms of a basis function (i.e., the wavelet) at different levels of resolution, resulting in a high-dimensional frequency-time representation. Lucas et al. [31], however, have demonstrated that for sEMG-based classification it is sufficient to preserve only the marginals of each level of the decomposition, thereby ignoring the time components of the decomposition and drastically reducing the dimensionality of the feature representation. In the following, this variant will be referred to as marginal Discrete Wavelet Transform (mDWT).

[5]Due to rectification onboard the Otto Bock electrodes, the MAV features are in our case closely related to RMS features.

TABLE II

	Method	Configuration
Feature Types	Mean Absolute Value (MAV)	
	Variance (VAR)	
	Waveform Length (WL)	
	sEMG Histogram (HIST)	10 bins, log. scale
	Cepstral Coefficients (CC)	first 5 coefficients
	Short-Time Fourier Transform (STFT)	4-sample rect. window
	marginal Discrete Wavelet Transform (mDWT)	sym4 wavelet, 3 levels
Classifiers	Support Vector Machine (SVM)	RBF kernel
	Multi-Layer Perceptron (MLP)	1 hidden layer, sigmoid
	k-Nearest Neighbors (k-NN)	
	Support Vector Machine (SVM)	linear kernel
	Linear Discriminant Analysis (LDA)	

In contrast to feature extraction methods, only a relatively limited set of classification methods have been employed for myoelectric movement classification. The classifiers considered here have all been used in related work and range from traditional statistical methods to more recent machine learning techniques. As simple methods we consider the well-known Linear Discriminant Analysis (LDA), k-Nearest Neighbors (k-NN), and a linear variant of the more recent Support Vector Machine (SVM). These methods are in contrast to the more powerful, non-linear classifiers. We consider the two most popular non-linear methods, namely the Multi-Layer Perceptron (MLP) (i.e., an artificial neural network) and the SVM with a Radial Basis Function (RBF) kernel. The motivation for considering both linear and non-linear methods is to verify whether the additional capacity of non-linear classifiers is in fact required to obtain satisfactory performance. A listing of the feature types and the classifiers and their configuration is given in Table II. Further details of the experimental setup can be found in the work of Kuzborskij et al. [32].

In accordance with the classification strategy by Englehart and Hudgins [33], the filtered signals are segmented into windows, for which we consider an increment of 10 ms (i.e., 1 sample) and windows lengths of 100 ms, 200 ms, and 400 ms. Five movement repetitions are used to train the classifier (subsampled at a regular interval of 10 samples), while all samples of the remaining five repetitions form the test set.

2) Results: Figure 4 presents the classification accuracies for all feasible combinations of feature type, classifier, and window lengths. Interestingly, several feature representations achieve a similar accuracy of around 76%, indicating that simpler features as MAV do not necessarily perform worse than advanced variants as mDWT. Furthermore, the non-linear SVM and MLP classifiers achieve similar maximum performance given an appropriate feature representation, although only SVM consistently achieves high performance when combined with most of the feature representations. The linear classifiers, on the other hand, perform poorly. Even though LDA has been found adequate for small-scale posture classification [34], our results demonstrates that linear classifiers in fact do not scale to a large number of postures. Finally, the optimal window length is dependent on both the classifier and the feature representation, although in the majority of cases the longer window length of 400 ms is preferable.

The classification accuracies in Figure 4 are encouraging, considering the large number of movements. An accuracy of 76%, however, would almost surely not be acceptable from the perspective of an actual end-user. Nonetheless, the scalar classification score obfuscates the fact that misclassifications are not evenly distributed over the duration of the movement. Figure 5, which relates classification errors with the time normalized for movement duration, demonstrates that misclassifications are primarily concentrated during the movement onset and offset. This is not surprising, since movements are continuous trajectories that transition gradually from one to another, in contrast to the abrupt changes of the discrete movement labels. Consequently, a drop in accuracy occurs primarily during these transitory periods, since the the change in movement is not yet clearly evident from the input sEMG signals.

This issue is illustrated in Figure 6, which visualizes ten trials of four postures (i.e., those used by Castellini et al. [13] plus the rest posture) for a single subject in the first two principal components[6]. When concentrating solely on the center of the movements (indicated by markers), then the three postures appear reasonably well separated. However, the trajectories overlap significantly on the transition from rest to movement and vice versa, causing a reduction of separability and hence misclassifications. This issue is relevant since (1) it demonstrates that accuracy is best improved by better distinguishing rest from movements during transitory phases, and (2) some related studies enforce separability by solely considering the center segment of the movement trajectory. A consequence of the latter is that these studies report overly optimistic classification results.

B. Statistical Analysis

The scalar average classification accuracy obfuscates how the accuracy is distributed over either subjects or movements.

[6]More precisely, these postures are extracted using Principal Component Analysis (PCA) over the entire dataset when using MAV features with a window length of 200 ms.

TABLE III
AVERAGED CLASSIFICATION ACCURACY WITH RESPECT TO SUBJECT PROPERTIES.

Property	Group	# Subj.	Accuracy
	All	27	0.7401 ± 0.0394
Gender	Female	7	0.7532 ± 0.0215
	Male	20	0.7356 ± 0.0431
Height	< 172 cm	10	0.7537 ± 0.0207
	172 to 180 cm	8	0.7497 ± 0.0422
	≥ 180 cm	9	0.7167 ± 0.0421
Weight	< 65 kg	9	0.7528 ± 0.0184
	65 to 75 kg	9	0.7512 ± 0.0416
	≥ 75 kg	9	0.7164 ± 0.0418
Age	< 27 y	7	0.7489 ± 0.0228
	27 to 29 y	10	0.7318 ± 0.0488
	≥ 29 y	10	0.7424 ± 0.0363

It is however feasible that certain subjects perform considerably worse than others (e.g., see the observations in [35]), or that certain movements are harder to discriminate than orders. Figure 7 demonstrates the distribution of classification accuracy over either subjects and movements. In order to eliminate sensitivity to a particular classifier or feature extraction method, the reported accuracy is the average accuracy over all combinations of the SVM and MLP classifiers with MAV, mDWT, HIST, and WL features based on window lengths of 100 ms, 200 ms, and 400 ms. All these combinations were found to perform similarly (see Figure 4). In case of subjects, there are no apparent outliers and the distribution does not significantly deviate from normality ($p = 0.338$, Shapiro-Wilk test). When considering the distribution over movements in Figure 7b, on the other hand, we observe a single outlier with very high performance, which corresponds to the rest posture. While the onset and offset of non-rest movements are often misclassified as rest, the rest posture itself is in fact nearly always correctly classified. This is aided by the fact that rest posture accounts for nearly 60% of all samples, causing the classifiers to be biased towards correctly classifying this specific class.

Even though the distribution over subjects is statistically not distinguishable from a normal distribution, this does not necessarily imply that all subjects are random samples from a single probability distribution. In contrast, it is likely that certain properties of the subjects affect classification accuracy. Such a relation between a subject's characteristics and classification accuracy is relevant in a clinical setting, as it helps to anticipate the rate of success or satisfaction of a prospective user of an active prosthesis. Table III lists the average accuracy for the total set of all subjects and for the subsets based on the properties asked in the questionnaire.

The results in Table III allow for several interesting observations, namely that classification accuracy is higher for female participants and that the accuracy decreases considerably with both subject height and weight. These observations, however, are strongly correlated, since male subjects are commonly taller and taller subjects are typically also heavier. This correlation can be largely eliminated by aggregating the height and

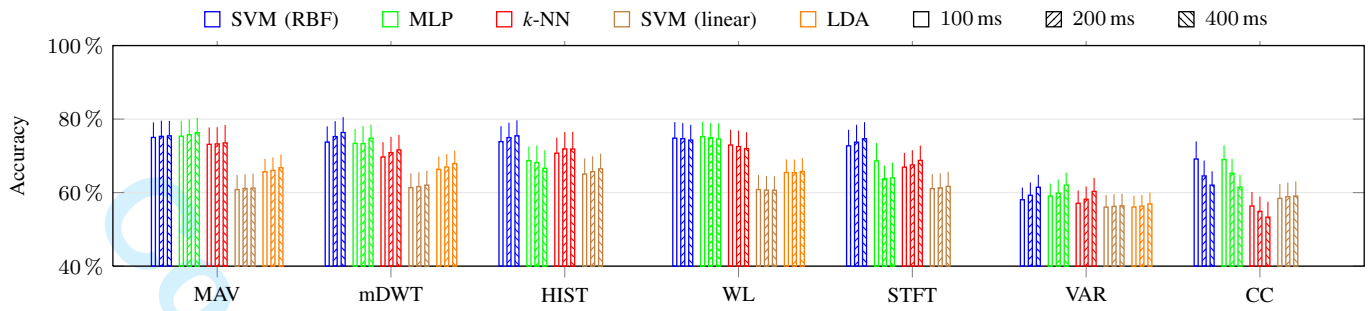


Fig. 4. Classification accuracies. Each bar represents method classification accuracy with respect to feature representation and window length, while line atop the bar is one standard deviation of accuracy. Classifiers are grouped by feature representations and labeled by different colors. Window lengths are represented in increasing order, namely 100ms, 200ms and 400ms and are tagged with different textures. LDA results are missing in case of the high dimensional STFT, CC, and HIST features due to non-singularity of the covariance matrix.

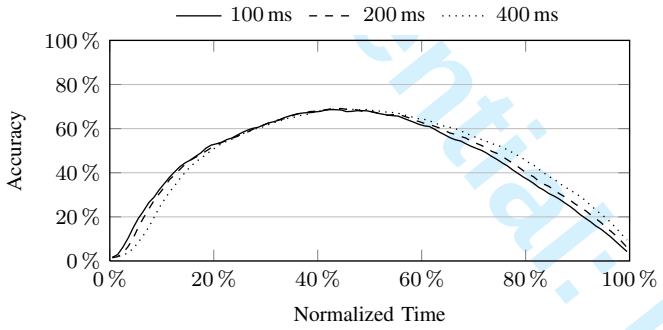


Fig. 5. Classification accuracy with respect to normalized movement duration for kernel SVM with MAV features and windows of 100, 200 and 400 ms. This figure is representative for other combinations of classifiers and feature types.

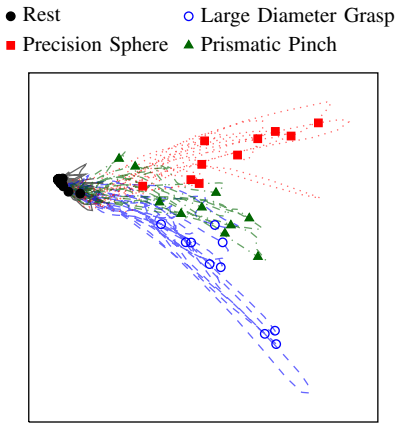


Fig. 6. Trajectories within the two principal components of all ten trials for three movements and the rest posture taken from a single subject. The samples in the temporal center of the trajectory are indicated with markers.

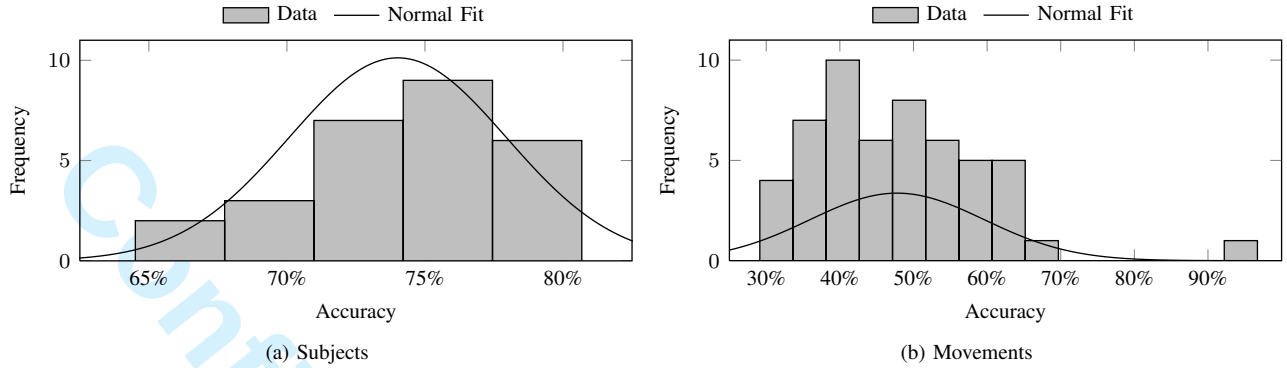
weight properties into the well-known BMI, which also limits the correlation with gender. Multiple regression analysis using the subject gender, age, and BMI as independent variables confirms that classification accuracy indeed decreases significantly with increasing BMI ($p = 0.021$, Student's t test), as demonstrated in Figure 8. On the other hand, the accuracy was not found to depend significantly on either gender ($p = 0.985$) or age ($p = 0.110$).

The negative relation between BMI and classification accuracy should not come as a surprise, since it is known that the adipose layer in the skin acts as an insulator [36, Chapter 3]. As a result, the amplitude and signal-to-noise ratio of the sEMG signal decrease, while cross-talk between muscles increases [37]. Both effects deteriorate the quality of the sEMG signal. To the best of our knowledge, our analysis is the first empirical confirmation that this signal deterioration indeed leads to significantly worse classification accuracy.

V. Conclusions and Discussion

This paper describes the release of the NINAPRO database, which aims to form a standard benchmarking resource for the biorobotics community. According to our knowledge, at the state of the art this database is the public sEMG database that includes more hand movements. The NINAPRO database consists of muscular activity gathered in controlled conditions using Otto Bock sEMG electrodes and kinematic data gathered using a CyberGlove and an inclinometer. Particular care is taken as far as electrode placement, device calibration, and data acquisition and synchronization are concerned. So far data is available for 27 intact subjects performing 10 successive repetitions of 52 hand, wrist, and forearm movements of interest. These movements have been selected via a careful examination of the literature and standard rehabilitation guidelines. The timings, repetitions, and durations of the stimuli were verified. The stimuli themselves are instructed using short movies that the subjects are asked to imitate. This makes the protocol extremely simple, stress-, and fatigue-free for the subjects.

A benchmark evaluation using a variety of popular feature extraction and classification methods in a continuous prediction setting reveals that the best performing methods achieve an accuracy of around 76%. In contrast to some related work, we found that the non-linear SVM and MLP classifiers perform considerably better than the linear SVM and LDA. Furthermore, the SVM with RBF kernel is to be preferred over the MLP classifier, as it showed similarly high performance for five out of seven feature represtations. This result also implies that relatively simple features as MAV can perform just as



(a) Subjects (b) Movements

Fig. 7. Histogram of the classification accuracies over (a) subjects and (b) movements. The solid line indicates a normal distribution fitted to the data.

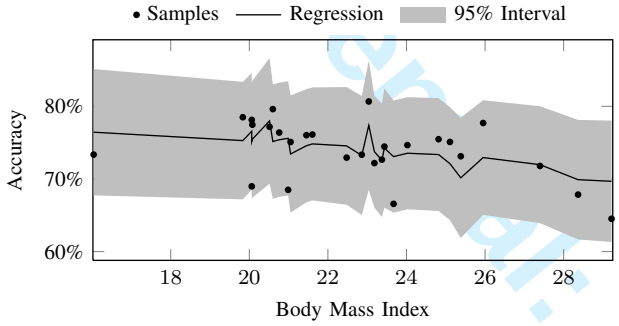


Fig. 8. Classification error versus the BMI for all 27 subjects as well as the predictions by the linear regressor. The small variations in the predictions are due to variations in age and gender among the subjects.

well as more advanced mDWT or STFT features, provided that they are combined with an appropriate classifier.

Further investigation demonstrates that misclassifications occur primarily during the movement onset and offset. The explanation for this phenomenon is that the sEMG signals are not yet (or not anymore) sufficiently discriminative in these transitory phases between movement and rest. This result implies that studies that ignore these phases are likely to overestimate true performance, and that efforts on improving movement classification in a realistic settings are best directed towards these ambiguous phases of the movements.

Multiple regression analysis of the classification accuracy with respect to several subject properties indicates that accuracy decreases significantly with an increasing BMI of the subject. While earlier studies have demonstrated that the adipose layer (estimated here using BMI) has a negative impact on the quality of the sEMG signal, our result is to the best of our knowledge the first empirical confirmation that this also affects movement classification accuracy. In contrast, the accuracy was not found to depend significantly either on subject gender or age.

A. Discussion and future work

Improvements to the setup: The classification results in Section IV-A indicated that more advanced feature representations as mDWT did not yield improvements over the simpler MAV or WL features. One possible explanation is that the rectification step onboard the Otto Bock electrodes is removing information that could potentially be exploited by advanced feature extraction methods. To eliminate this possibility, we are migrating the acquisition setup to a set of Delsys™ Trigno Wireless® electrodes. As opposed to the rectified and filtered signals from the Otto Bock electrodes, the Trigno electrodes return the raw sEMG signal at 2 kHz sampling rate. Furthermore, these electrodes are wireless (thus less restrictive for subjects) and also contain a 3-axis accelerometer. The latter property will allow us to investigate to which extent accelerometry can aid movement classification.

Second, it is of interest to also gather *force* data while performing the actions of interest, rather than kinematic data only. This has a double motivation: (a) sEMG can be naturally associated with *graded forces* as well as with movements and postures (see, e.g., [2, 38]); (b) the use of regression rather than classification can dramatically increase the dexterity of the control, shifting from a finite set of predetermined postures to an infinite manifold of hand configurations. We plan to employ the Finger-Force Linear Sensor (FFLS), a synergistic finger-force measurement device [39]. Adaptation and calibration of the new devices is already done, and we plan to include data obtained using them in the next database release.

Further considerations: The database described in this paper should be regarded a first milestone toward the larger goal of providing the biorobotics community with a large-scale database for use in hand prosthetics research. The experience obtained with the acquisition protocol and setup using intact subjects will prove useful in the next phase of acquisitions from amputated persons. Posture recognition from amputees is typically more difficult, since muscle activity decreases due to lack of use and since muscles may be damaged due to trauma or surgical intervention. The availability of data recorded from amputees is however crucial to perform experiments on movement classification that accurately reflects real-world conditions.

ACKNOWLEDGMENT

The authors would like to thank all the subjects that participated in the data acquisitions. This work is partially supported by the Swiss National Science Foundation Sinergia project NINAPRO.

REFERENCES

[1] F. R. Finley and R. W. Wirta, "Myocoder studies of multiple myopotential response," *Archives of Physical Medicine and Rehabilitation*, vol. 48, no. 11, pp. 598–601, 1967.

[2] C. J. D. Luca, "The use of surface electromyography in biomechanics," *Journal of Applied Biomechanics*, vol. 13, no. 2, pp. 135–163, 1997.

[3] R. Merletti, A. Botter, A. Troiano, E. Merlo, and M. Minetto, "Technology and instrumentation for detection and conditioning of the surface electromyographic signal: State of the art," *Clinical Biomechanics*, vol. 24, pp. 122–134, 2009.

[4] D. J. Atkins, "Epidemiologic overview of individuals with upper-limb loss and their reported research priorities," *Journal of Prosthetics & Orthotics*, vol. 8, no. 1, pp. 2–11, 1996.

[5] S. Micera, J. Carpaneto, and S. Raspopovic, "Control of hand prostheses using peripheral information," *IEEE Reviews in Biomedical Engineering*, vol. 3, pp. 48–68, Octobre 2010.

[6] M. Everingham, L. V. Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes Challenge 2010 (VOC2010) Results," http://www.pascal-network.org/challenges/VOC/voc2010/workshop/index.html, 2010.

[7] G. Griffin, A. Holub, and P. Perona, "Caltech-256 object category dataset," California Institute of Technology, Tech. Rep. 7694, 2007. [Online]. Available: http://authors.library.caltech.edu/7694

[8] J. Xiao, J. Hays, K. Ehinger, A. Oliva, and A. Torralba, "Sun database: Large-scale scene recognition from abbey to zoo," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, 2010, pp. 3485–3492.

[9] A. Howard and N. Roy, "The robotics data set repository (radish)," 2003. [Online]. Available: http://radish.sourceforge.net/

[10] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of rgb-d slam systems," in *Proceedings of the International Conference on Intelligent Robot Systems (IROS)*, 10 2012.

[11] H. Müller, A. Rosset, J.-P. Vallée, F. Terrier, and A. Geissbuhler, "A reference data set for the evaluation of medical image retrieval systems," *Computerized Medical Imaging and Graphics*, vol. 28, no. 6, pp. 295–305, 2004.

[12] M. Atzori, A. Gijsberts, S. Heynen, A.-G. Mittaz-Hager, O. Deriaz, P. van der Smagt, C. Castellini, B. Caputo, and H. Müller, "Building the NINAPRO database: A resource for the biorobotics community," in *Proceedings of the IEEE International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, 2012, pp. 1258–1265.

[13] C. Castellini, A. E. Fiorilla, and G. Sandini, "Multi-subject / daily-life activity EMG-based control of mechanical hands," *Journal of Neuroengineering and Rehabilitation*, vol. 6, no. 41, 2009.

[14] F. V. Tenore, A. Ramos, A. Fahmy, S. Acharya, R. Etienne-Cummings, and N. V. Thakor, "Decoding of individuated finger movements using surface electromyography," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 5, pp. 1427–1434, 2009.

[15] H. Tsuji, H. Ichinobe, K. Ito, and M. Nagamachi, "Discrimination of forearm motions from emg signals by error back propagation typed neural network using entropy," *IEEE Transactions, Society of Instrument and Control Engineers*, vol. 29, no. 10, pp. 1213–1220, 1993.

[16] 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, no. 2, pp. 210–222, April 2003.

[17] C. Castellini, E. Gruppioni, A. Davalli, and G. Sandini, "Fine detection of grasp force and posture by amputees via surface electromyography," *Journal of Physiology (Paris)*, vol. 103, no. 3-5, pp. 255–262, 2009.

[18] M. Atzori, C. Castellini, and H. Müller, "Spatial registration of hand muscle electromyography signals," in *7th International Workshop on Biosignal Interpretation*, Como, July 2012.

[19] T. Feix, "Grasp taxonomy comparison," Otto Bock GmbH, Tech. Rep., 2008. [Online]. Available: http://grasp.xief.net

[20] M. R. Cutkosky, "On grasp choice, grasp models, and the design of hands for manufacturing tasks," *IEEE Transactions on Robotics and Automation*, vol. 5, no. 3, pp. 269–279, June 1989.

[21] N. Kamakura, M. Matsuo, H. Ishii, F. Mitsuboshi, and Y. Miura, "Patterns of static prehension in normal hands," *The American journal of occupational therapy: official publication of the American Occupational Therapy Association*, vol. 34, no. 7, pp. 437–445, 1980.

[22] S. J. Edwards, D. J. Buckland, and J. D. McCoy-Powlen, *Developmental and Functional Hand Grasps*. Slack Incorporated, 2002.

[23] P. L. Hudak, P. C. Amadio, and C. Bombardier, "Development of an upper extremity outcome measure: the dash (disabilities of the arm, shoulder and hand)," *American Journal of Industrial Medicine*, vol. 29, no. 6, pp. 602–608, 1996.

[24] R. Kato, H. Yokoi, and T. Arai, "Competitive learning method for robust emg-to-motion classifier," in *Proceedings Intelligent Autonomus Systems*, 2006, pp. 946–953.

[25] F. C. P. Sebelius, B. N. Rosen, and G. N. Lundborg, "Refined myoelectric control in below-elbow amputees using artificial neural networks and a data glove," *Journal of Hand Surgery*, vol. 30, no. 4, pp. 780–789, July 2005.

[26] T. R. Farrell and R. F. Weir, "A comparison of the effects of electrode implantation and targeting on pattern classification accuracy for prosthesis control," *IEEE Transactions on Biomedical Engineering*, vol. 55, pp. 2198–2211, March 2008.

[27] B. Crawford, K. Miller, P. Shenoy, and R. Rao, "Real-time classification of electromyographic signals for robotic control," in *Proceedings of AAAI*, 2005, pp. 523–528.

[28] G. Staude, "Objective motor response onset detection in surface myoelectric signals," *Medical Engineering & Physics*, vol. 21, no. 6-7, pp. 449–467, 1999.
[29] J. D. Hamilton, *Time Series Analysis*. Princeton University Press, 1994.
[30] M. Zecca, S. Micera, M. C. Carrozza, and P. Dario, "Control of multifunctional prosthetic hands by processing the electromyographic signal," *Critical Reviews in Biomedical Engineering*, vol. 30, no. 4-6, pp. 459–485, 2002.
[31] M. 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.
[32] I. Kuzborskij, A. Gijsberts, and B. Caputo, "On the challenge of classifying 52 hand movements from surface electromyography," in *Proceedings of EMBC - the 34th annual conference of the IEEE Engineering in Medicine and Biology Society*, 2012, pp. 4931–4937.
[33] K. Englehart and B. Hudgins, "A robust, real-time control scheme for multifunction myoelectric control," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 7, pp. 848–854, 2003.
[34] 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, 7 1999.
[35] H. Bouwsema, C. K. van der Sluis, and R. M. Bongers, "Learning to control opening and closing a myoelectric hand," *Archives of Physical Medicine and Rehabilitation*, vol. 91, no. 9, pp. 1442–1446, 2010.
[36] E. Criswell, *Cram's Introduction to Surface Electromyography*. Jones & Bartlett Learning, 2010.
[37] T. A. Kuiken, M. M. Lowery, and N. S. Stoykov, "The effect of subcutaneous fat on myoelectric signal amplitude and cross-talk." *Prosthetics and Orthotics International*, vol. 27, no. 1, pp. 48–54, 2003.
[38] C. Castellini and R. Kõiva, "Using surface electromyography to predict single finger forces," in *Proceedings of BioRob - IEEE International Conference on Biomedical Robotics and Biomechatronics*, 2012, pp. 1266–1272.
[39] R. Kõiva, B. Hilsenbeck, and C. Castellini, "FFLS: An accurate linear device for measuring synergistic finger contractions," in *Proceedings of EMBC - the 34th annual conference of the IEEE Engineering in Medicine and Biology Society*, 2012, pp. 531–534.