Characterization of a Benchmark Database for Myoelectric Movement Classification

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

Abstract—In this paper, we characterize the NINAPRO database and its use as a benchmark for hand prosthesis evaluation. The database is a publicly available resource that aims to support research on advanced myoelectric hand prostheses. 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. Besides describing the acquisition protocol, overall features of the datasets and the processing procedures in detail, we present benchmark classification results using a variety of feature representations and classifiers. Our comparison shows that simple feature representations such as mean absolute value and waveform length can achieve similar performance to the computationally more demanding marginal discrete wavelet transform. With respect to classification methods, the nonlinear support vector machine was found to be the only method consistently achieving high performance regardless of the type of feature representation. Furthermore, statistical analysis of these results shows that classification accuracy is negatively correlated with the subject's Body Mass Index. The analysis and the results described in this paper aim to be a strong baseline for the NINAPRO database. Thanks to the NINAPRO database (and the characterization described in this paper), the scientific community has the opportunity to converge to a common position on hand movement recognition by surface electromyography, a field capable to strongly affect hand prosthesis capabilities.

Index Terms—Electromyography, machine learning, prosthetics, publicly available databases.

I. INTRODUCTION

SINCE the 1960s, pattern recognition algorithms have been applied on surface electromyography (sEMG) signals to control simple prosthetic grippers with a single degree of

freedom (DOF) [1], [2]. 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. An amputee should be able to dexterously control the prosthesis just by desiring to do so in a natural way. Still, 45 years later this goal has not yet been reached, since one quarter to one third of the amputees reject self-powered prostheses. Low reliability, trouble with maintenance, and low dexterity are some of the problems that lead to this situation [3], [4].

A major obstacle towards the goal of natural prosthesis use 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 there standardized databases available for research. 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 [5], CALTECH 256 [6], SUN [7]), robotics (e.g., Radish [8], RGB-D SLAM [9]), medical information retrieval (e.g., ImageCLEF [10]), 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.

First, we present the first Non-Invasive Adaptive Prosthetics (NINAPRO) database in detail, which was succinctly introduced by [11]. 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
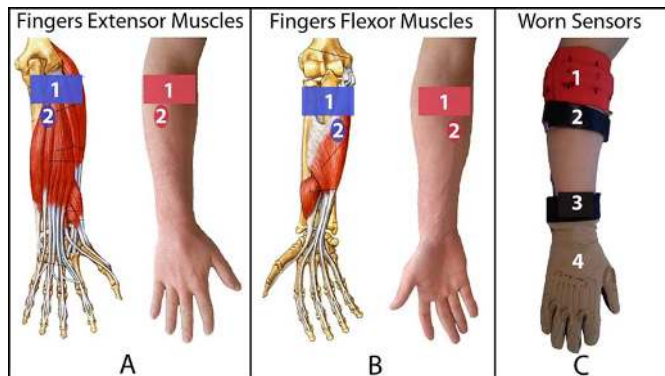
Fig. 1. Placement of the electrodes. A: sEMG electrodes placed on finger extensor muscles (A1: equally spaced electrodes; A2: spare electrode). B: sEMG electrodes placed on finger flexor muscles (B1: equally spaced electrodes; B2: spare electrode). C: All the sensors positioned on the arm (C1: equally spaced electrodes; C2: Spare electrode; C3: Inclinometer; C4: CyberGlove II).

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 also in the future.

Second, we characterize the NINAPRO database, with overall characteristics of the datasets and with an extensive benchmark comparison using a large variety of popular feature extraction and classification methods. This benchmark evaluation is intended as a strong reference that facilitates and gives direction to future experiments on the database. Nonetheless, the evaluation is also useful in its own right, since it allows to assess whether popular approaches are able to scale to 52 movement classes. For instance, the computationally simple linear discriminant analysis has shown excellent performance in a number of studies [12]–[14]. Whether this method performs equally well in a large-scale setting is an open question; more advanced nonlinear classifiers may be required to cope with the increased problem complexity. Finally, we apply statistical analysis to determine which of a subject's properties affect classification accuracy and how to potentially take these additional parameters into account. Insight on these points are relevant from a clinical point of view, as they provide an *a priori* estimate on the success or even acceptance of a myoelectric prosthesis for a specific patient.

The analysis and the results described in this paper are an example of what can be done with the NINAPRO database and they are aimed to lay the foundations for comparative analyses by research groups interested in the topic.

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 a principled relabeling strategy to correct erroneous labels. Section IV presents a description of overall features of the dataset, a benchmark evaluation on the database and 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 described in this paper 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 10 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 5. Particular care was taken in deciding the *placement* of the electrodes on the forearm. 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], [15] as well as on amputees [16]. However, early research on pattern recognition for sEMG [17], [18] (recently confirmed in [19]) 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 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 Fig. 1). This positioning of the electrodes also gives the opportunity to improve classification results by applying linear and nonlinear spatial registration algorithms, as described in Atzori *et al.* [20].

2) Kinematics: The kinematic configuration of the hand is measured using a 22-sensor CyberGlove II dataglove[2], shown in the right panel of Fig. 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 1° depending on the size of the subject's hand, carefully wearing the glove and the angular range of the considered joint. We record raw sensor values rather than estimated joint angles, the reason being that reliable calibration of the glove is prohibitively time-consuming. Most machine learning techniques are invariant to linear scaling of the data and calibration is in these cases unnecessary; moreover this decision gives more data processing freedom to the final users of the data. If desired, exact joint angles can be obtained by calibrating the glove a posteriori for a given subject.

In addition to the CyberGlove, a standard commercially available two-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) Signal 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 provided per device is associated with an accurate timestamp (using Windows Performance Counters) and directly written to mass storage.

B. Data Acquisition

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. 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).

In the case of amputees, we also note the date, 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 and DASH (disability of the arm, shoulder, and hand) score [21]. Moreover, we take pictures of the arm with and without the acquisition setup to permit *a posteriori* quality control.

After finalizing the forms, the subject is asked to sit comfortably on an adjustable chair in front of a table with a large monitor. Intact subjects wear the sEMG electrodes, dataglove, and inclinometer on the right arm. Amputees wear the sEMG electrodes on the amputated arm while the dataglove and the inclinometer are worn on the intact limb. While intact subjects execute the experiments with a monolateral imitation method (using the right hand), amputated subjects use a bilateral imitation procedure. This difference is determined by the fact that amputees can in principle not provide ground truth, because the absence of the limb makes it impossible to gather kinematic and/or force data from it. Therefore, there is an ongoing debate in the community about how to get a sensible ground truth in this case. In general, monolateral visual imitation and bilateral imitation have been proposed [19], [22]. We decided to choose monolateral imitation for the intact subjects (which permits to get a natural repetition of the movements) and bilateral imitation for amputated subjects (which permits to obtain a ground truth) in order to maximize the ground truth obtained by both groups. The imitation stimulus is visual. A movie of the movement that the subject repeats is displayed on the screen of a laptop and the subject is asked to repeat it as naturally as possible.

First, the subjects have to perform a training sequence that involves three repetitions of a selection of movements in order to get accustomed to the protocol. Then, the real data acquisition starts and the subjects have to repeat the 52 movements 10 times. The number of repetitions was chosen in order to be consistent with the state of the art [23], [24], [19] and to avoid fatigue in the subjects. Each movement repetition lasts 5 s and is followed by 3 s of rest. The movements were selected from the hand taxonomy and robotics literature (see, e.g., [25]–[28]). The experiment is divided into three exercises (Table I and Fig. 2).

1) 12 basic movements of the fingers (flexions and extensions).

TABLE I
SYNTHETIC DESCRIPTIONS OF THE 52 MOVEMENTS INCLUDED IN THE ACQUISITION PROTOCOL, ALONG WITH A REFERENCE, IF AVAILABLE

	#	Description	Ref.
Finger mvts.	1-2	Index flexion and extension	[16]
	3-4	Middle flexion and extension	[16]
	5-6	Ring flexion and extension	[16]
	7-8	Little finger flexion and extension	[16]
	9-10	Thumb adduction and abduction	[16]
	11-12	Thumb flexion and extension	
Hand postures	1	Thumb up	[29]
	2	Extension of index and middle finger while flexing others (cf. "V-sign")	
	3	Flexion of ring and little finger while extending others	[23]
	4	Thumb opposing base of little finger	[23]
	5	Abduction of the fingers	[23]
	6	Fingers flexed together in fist	[23]
	7	Pointing index	[24]
	8	Adduction of extended fingers	[30]
Wrist mvts.	1-2	Wrist supination and pronation (rotation axis through the middle finger)	[29]
	3-4	Wrist supination and pronation (rotation axis through the little finger)	
	5-6	Wrist flexion and extension	[29]
	7-8	Wrist radial and ulnar deviation	[30]
	9	Wrist extension with closed hand	
Grasping and functional movements	1-2	Large and small diameter grasp	[25]
	3	Fixed hook grasp	[25]
	4	Index finger extension grasp	[25]
	5	Medium wrap	[25]
	6	Ring grasp	[25]
	7	Prismatic four fingers grasp	[25]
	8	Stick grasp	[25]
	9	Writing tripod grasp	[25]
	10-12	Power, three finger, and precision sphere grasp	[25]
	13	Tripod grasp	[25]
	14-15	Prismatic and tip pinch grasp	[25]
	16	Quadpod grasp	[25]
	17	Lateral grasp	[25]
	18	Parallel extension grasp	[25]
	19	Extension type grasp	[25]
	20	Power disk grasp	[25]
	21	Open a bottle with a tripod grasp	
	22	Turn a screw (grasp the screwdriver with a stick grasp)	
	23	Cut something (grasp the knife with an index finger extension grasp)	

2) Eight isometric and isotonic hand configurations and nine basic movements of the wrist (adduction/abduction, flexion/extension, and pronation/supination).

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.

The exercises last respectively 16, 23, and 31 min. Subjects are allowed short breaks between exercises to avoid muscle fatigue, such that the total duration of the experiment is around 100 min (including preparatory steps and training). The sequence of movements is not randomized in order to induce
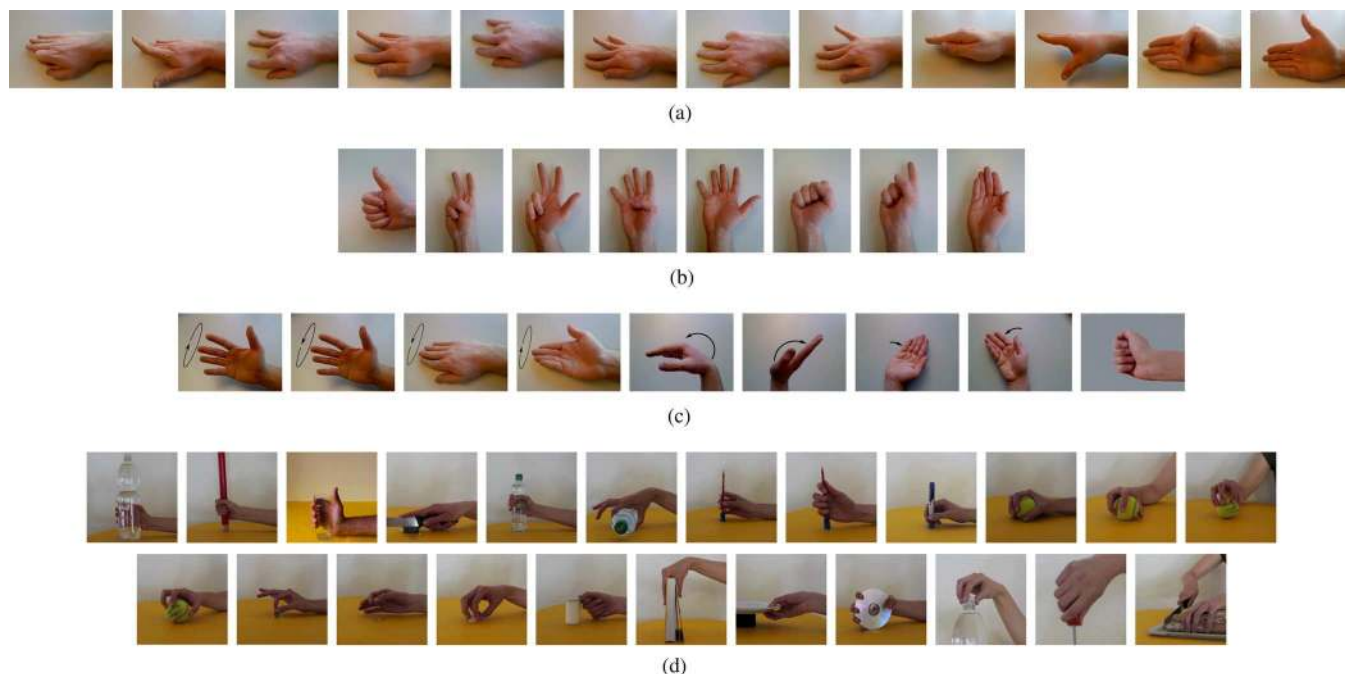
Fig. 2.   The 52 movements included in the acquisition protocol. (a) Basic movements of the fingers (flexions and extensions). (b) Isometric, isotonic hand config-urations ("hand postures"). (c) Basic movements of the wrist. (d) Grasping and functional movements.

unconscious movement repetitions into the subjects. The rep-etitions of each movement were performed in a block, and the order of the blocks was the same across participants. This setting permitted to reduce the number of movement repetition errors in comparison to randomized protocols. The experiment received the approval of Ethics Commission of the Canton of Valais (Switzerland), where all acquisitions were performed.

Finally, the data of each subject are stored anonymously in a database and made available on a website[4]. 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.

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 kine-matic and sEMG streams with the stimulus, removal of noise 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.

A. Synchronization and Filtering

Synchronization of the input modalities is relatively straight-forward with the NINAPRO database, since an accurate time-stamp is assigned to each data point with respect to a single ref-

[4]The database can be accessed at http://ninapro.hevs.ch. 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.

erence time based on the CPU's (invariant) timestamp counter. The difference in sampling rates is eliminated by linearly inter-polating 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.

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 mis-alignment is shown in the left panel of Fig. 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 [31], we first remove irrelevant autoregressive components by whitening the rectified signals using a multivariate $\mathrm{VAR}(p)$ model [32]. 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 Fig. 3, center)
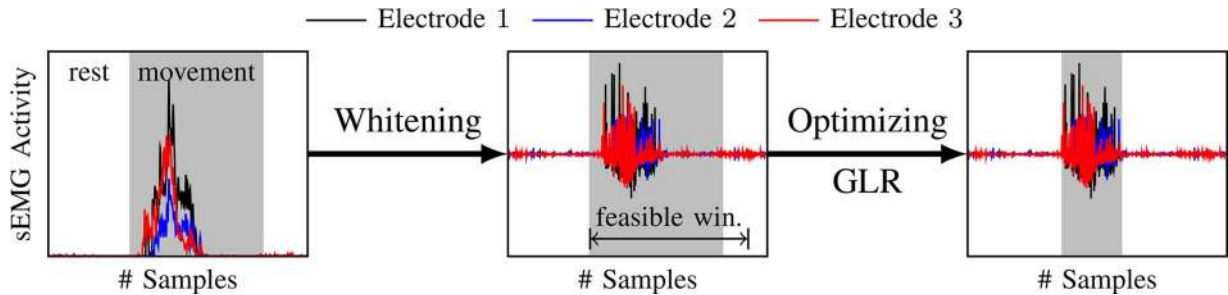
Fig. 3. Illustration of the relabeling process consisting of whitening the rectified sEMG signal and optimization of the GLR within the feasible window. Shaded area indicates time windows labeled as a nonrest movement. For clarity the figure displays only three out of 10 electrodes.

of length T is then divided into 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

$$\underset{1 \leq t_0 \leq T}{\arg\max} \ \underset{t_0 \leq t_1 \leq T}{\arg\max} \ \underset{\boldsymbol{\theta}_0 \in \Theta_0}{\sup} \ \underset{\boldsymbol{\theta}_1 \in \Theta_1}{\sup}$$

$$\left[\sum_{i=1}^{t_0-1} \ln p_{\boldsymbol{\theta}_0}(\mathbf{y}_i) + \sum_{j=t_0}^{t_1-1} \ln p_{\boldsymbol{\theta}_1}\mathbf{y}_j + \sum_{k=t_1}^{T} \ln p_{\boldsymbol{\theta}_0}(\mathbf{y}_k) \right]. \quad (1)$$

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

In order to characterize the database, we present an analysis of the variability of the signal and a classification benchmark obtained using various feature extraction and classification methods as a strong baseline. These results are not only intended as a reference for future work, but also to investigate whether the methods can actually scale to this large number of movement classes. Smith *et al.* [33] have found that both accurate identification of movements as well as low prediction delay are important factors that determine online controllability. To increase controllability and acceptance, it is therefore necessary to establish approaches that maximize movement classification

accuracy. Finally, the comparatively large number of subjects in the database allow us to subsequently determine the influence of the subjects' properties on classification accuracy.

A. Signal Variability

Many factors can affect the sEMG signal, including anatomical characteristics, fatigue and detection-system parameters [34], [35]. These factors contribute to create the recorded signal itself, and therefore they can influence classification accuracy. To ensure the quality of the NINAPRO data, we evaluate the variability of the signal across subjects, movements and movement repetitions. We consider both the mean amplitude of the sEMG signal as well as the mean amplitude of the data glove sensors (as an estimate of the hand motion range) (Figs. 4–6). No evident conclusions come from Fig. 4. The correlation between the average movement motion range and the sEMG amplitude (Fig. 5) is 0.73, in accordance with the relationship between the two variables. Finally, we observe a decreasing trend through the 10 repetitions of each movement in both the sEMG amplitude and in the hand motion range (Fig. 6), which is obfuscated by the high standard deviations of the data. Analyzing each subject separately by linear regression, a significant ($p < 0.05$) dependence of the signal amplitude by the repetition is obtained in 63% of cases considering the range of motion, and in 22% of cases considering the amplitude of the sEMG.

B. Classification Benchmark

Successful classification of movements from sEMG signals depends to a large extent on the type of feature representations and classifiers. To ensure that we establish a strong performance, we consider a variety of popular feature extraction and classification methods. Besides producing a direct quantitative comparison of these methods this benchmark also investigates whether current state-of-the-art methods can indeed reach satisfactory performance levels in 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 [36], [4]. 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
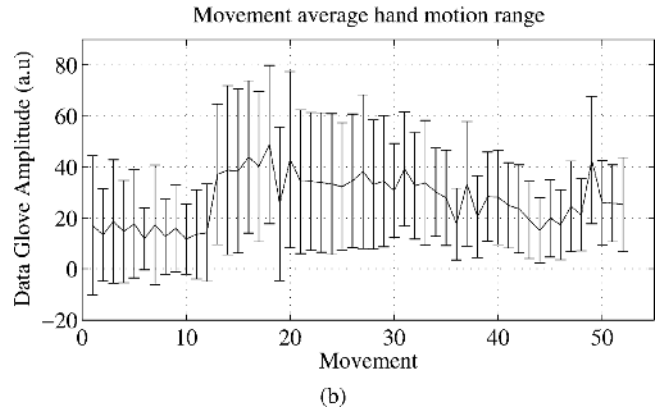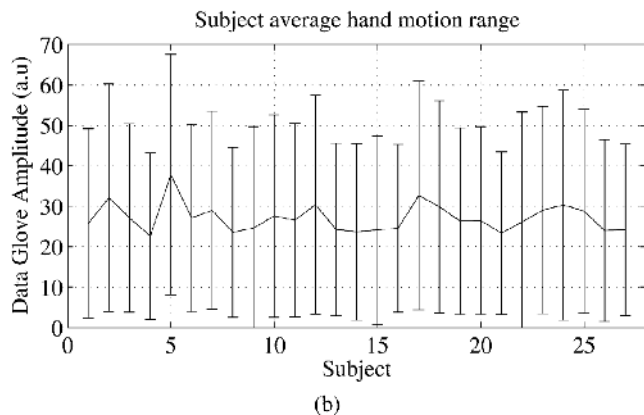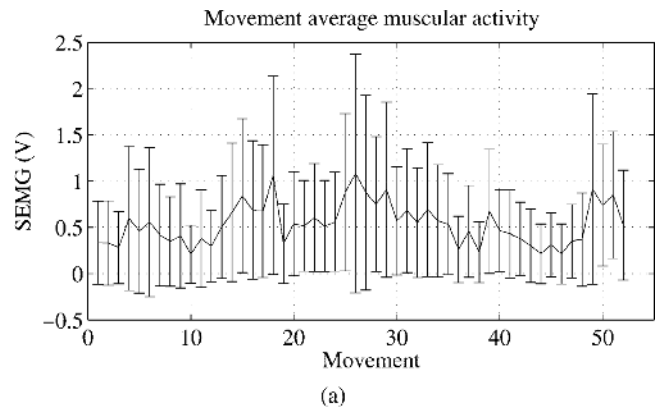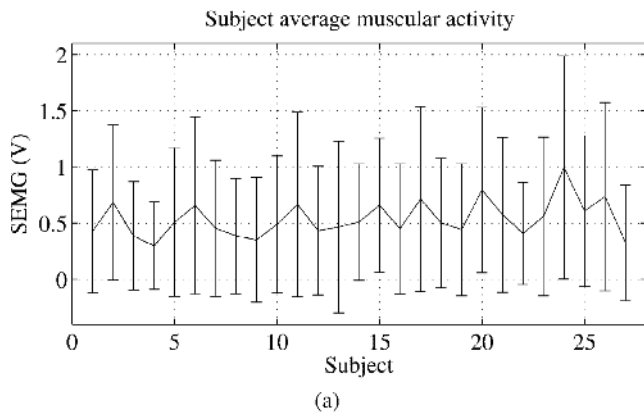
Fig. 4.   Variability across subjects of SEMG (a) and data glove (b) average amplitude.

Fig. 5.   Variability across movements of SEMG (a) and data glove (b) average amplitude.

of these simple time domain features we consider the 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 resolutions, resulting in a high-dimensional frequency-time representation Lucas *et al.* [37], 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 is referred to as marginal discrete wavelet transform (mDWT).

In contrast to feature extraction methods, only a relatively limited set of classification methods have been employed for
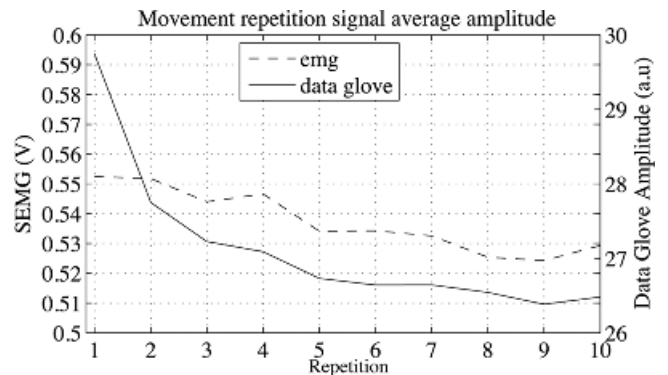
---

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

Fig. 6.   Variability of SEMG and data glove average amplitude across movement repetitions. *(The standard deviation of each data point is over the ordinate axis limits).*

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 more powerful, nonlinear classifiers. We consider the two most popular nonlinear 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 nonlinear methods is to verify whether the

TABLE II
FEATURE AND CLASSIFIER CONFIGURATIONS

	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)	

additional capacity of nonlinear classifiers is in fact required to obtain satisfactory performance. Table II shows a summary of the considered feature types and classifiers, and their respective configurations. Further details on the experimental setup can be found in the work of Kuzborskij *et al.* [38].

2) Data Processing and Split: In accordance with the classification strategy by Englehart and Hudgins [39], the filtered signals are segmented into a continuous stream of windows. We consider a sliding window with an increment of 10 ms (i.e., a window for every sample) and window lengths of 100, 200, and 400 ms . The features listed in Table II are then extracted from each individual window in this stream, while the movement label is defined as the movement type at the moment of the most recent sample within a window. The features extracted from the windows together with the corresponding movement labels then become samples for the classifier. The total number of the samples collected in this manner for a single subject is approximately $450\,000$.

This total set of classification samples is subsequently split into train and test sets based on movement repetitions; repetitions $\{1, 3, 4, 5, 9\}$ are used for training, while the repetitions $\{2, 6, 7, 8, 10\}$ are dedicated to testing. Furthermore, the training set is regularly subsampled at a factor of 10 for computational reasons. After this split and subsampling of the training set, we have roughly $25\,000$ training samples and $200\,000$ test samples. Since each movement repetition is preceded by a rest, these sets consist for (slightly more than) 50% of samples belonging to the rest posture. The duration of the 52 nonrest movements does not vary much (even after relabeling), so that each movement is responsible for almost 1% of the samples.

3) Results: Fig. 7 presents the classification accuracies on the complete set of 53 movements (the rest stae is included as additional "movement") for all feasible combinations of feature type, classifier, and window lengths. In our setting, the accuracy refers to the ratio of test samples (or windows) that were classified correctly. Interestingly, several feature representations achieve a similar accuracy of around 76%, indicating that simpler features such as MAV do not necessarily perform worse than advanced variants such as mDWT. Furthermore, the nonlinear 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

often been found adequate for small-scale posture classification [40], [12], [13], 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 the cases a longer window length of 400 ms is preferable.

The classification accuracies in Fig. 7 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. Fig. 8, 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 change in movement is not yet clearly evident from the input sEMG signals. Misclassifications during these phases are best characterized as a delay in predicting the correct class, rather than incorrectly classifying one movement as another. Besides classification accuracy, also prediction delays have been identified as an important factor that affects online controllability [33].

This is illustrated in Fig. 9, which visualizes 10 repetitions of a subset of four movements (i.e., those used by Castellini *et al.* [15] 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 they 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 is relevant since 1) it demonstrates that accuracy is best improved by better distinguishing rest from movements during the 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 overestimate classification accuracy.

C. Statistical Analysis

The scalar average classification accuracy obfuscates how the accuracy is distributed over both subjects and movements. It is however feasible that certain subjects perform considerably worse than others (e.g., see the observations in [41]), or that certain movements are harder to discriminate than others. Fig. 10 shows the distribution of the classification accuracy over 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, 200, and 400 ms. All combinations were found to perform similarly (see Fig. 7). In case of subjects, there are no apparent outliers and the distribution does not significantly deviate from normality ($p = $

---

[6]More precisely, these postures are extracted using (PCA) over the entire dataset when using MAV features with a window length of 200 ms.
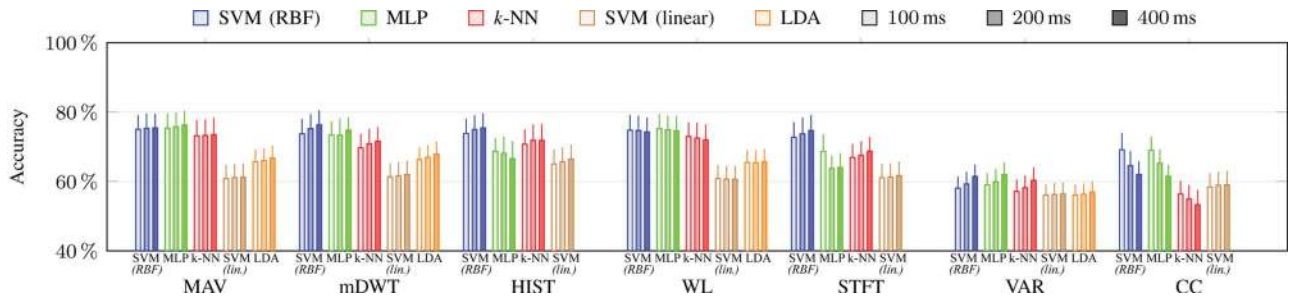
Fig. 7. Accuracies on the 53-class classification problem for combinations of learning method, feature representation and window length. Each bar represents method classification accuracy with respect to feature representation and window length, while lines atop the bar are one standard deviation of accuracy. Classifiers are grouped by feature representations and labeled by colors. Window lengths of 100, 200, and 400 ms are shown in increasing order and increasing opacity. LDA results are missing in case of the high dimensional STFT, CC, and HIST features due to nonsingularity of the covariance matrix.
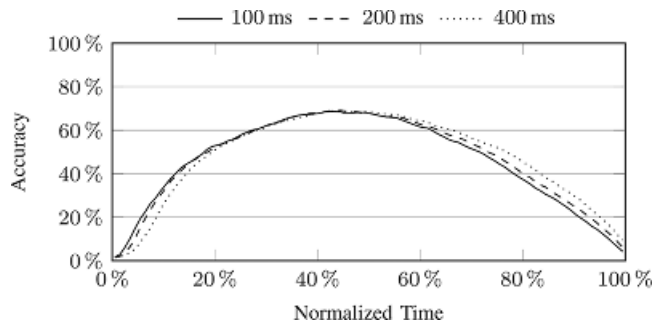

Fig. 8. Classification accuracy averaged over the 52 nonrest movements with respect to normalized movement duration for kernel SVM with MAV features and windows of 100, 200, and 400 ms. This figure is representative for other combinations of classifiers and feature types.
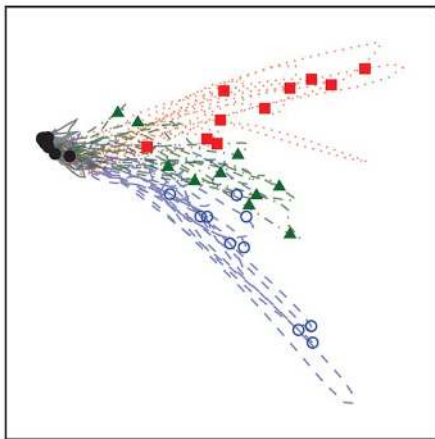

Fig. 9. Trajectories within the two principal components of all 10 repetitions for an example subset of three movements and the rest posture taken from a single subject. Samples in the temporal center of the trajectory are indicated with markers.

0.338, Shapiro–Wilk test). When considering the distribution over movements in Fig. 10(b), 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 nonrest movements are often misclassified as rest, the rest posture itself is in fact nearly always correctly classified. This is helped by the fact that rest posture accounts for more than 50% 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. It is likely that certain properties of the subjects affect classification gathered accuracy. Such a relation between a subject's characteristics and classification accuracy is relevant in a clinical setting, as it helps to anticipate the success rate 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 information in the questionnaire.

The results in Table III allow several interesting observations. Multiple regression analysis using the subject gender, laterality, age, and the well-known (BMI)

$$ \mathrm{BMI} = \frac{\mathrm{mass}_{\mathrm{kg}}}{(\mathrm{height}_{\mathrm{m}})^2}, $$

as independent variables confirms that classification accuracy decreases significantly with increasing BMI ($p = 0.026$, Student's t test), as demonstrated in Fig. 11. On the other hand, accuracy was not found to depend significantly on either gender ($p = 1$), laterality ($p = 0.921$), or age ($p = 0.117$).

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 [42, Ch. 3]. As a result, the amplitude and signal-to-noise ratio of the myoelectric signals decrease, while cross-talk between muscles increases [34]. Both effects decrease the quality of the signals. Our analysis suggests that this signal deterioration also leads to significantly worse classification accuracy.

V. CONCLUSION AND DISCUSSION

This paper describes the first NINAPRO database in detail and characterizes it with global analyses and benchmark classification results. The NINAPRO database aims at forming a standard benchmarking resource for the biorobotics community. At the moment of writing it is the only publicly available database relating sEMG to hand movements, it contains a much larger number of subjects and of hand movements with respect to related work. The dataset 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 was taken as far as electrode placement,
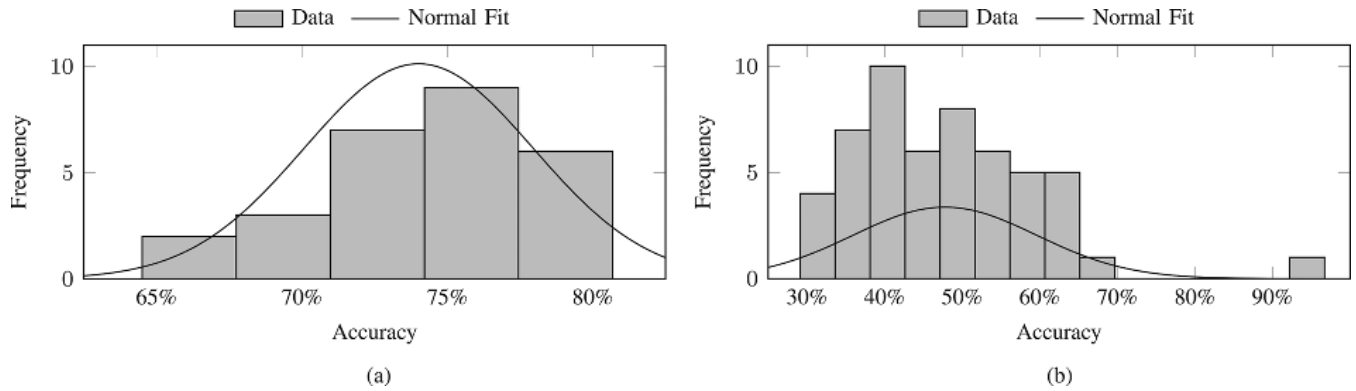
Fig. 10.   Histogram of the classification accuracies over (a) subjects and (b) movements. Solid line indicates a normal distribution fitted to the data.

TABLE III
AVERAGED CLASSIFICATION ACCURACY WITH RESPECT
TO SUBJECT PROPERTIES

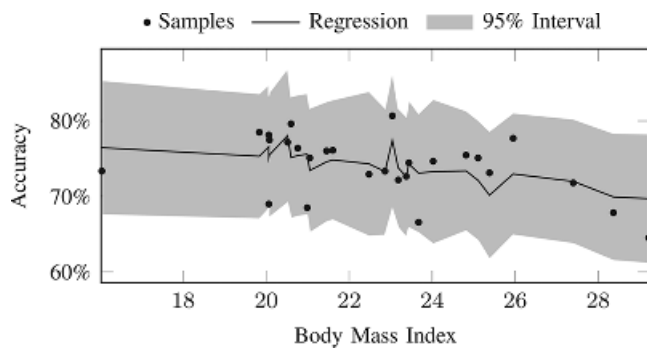| Property	Group	# Subj.	Accuracy
	All	27	0.7401 ± 0.0394
Gender	Female	7	0.7532 ± 0.0215
	Male	20	0.7356 ± 0.0431
Laterality	Left	2	0.7380 ± 0.0086
	Right	25	0.7403 ± 0.0409
Height	$< 172\,\mathrm{cm}$	10	0.7537 ± 0.0207
	172 to 180 cm	8	0.7497 ± 0.0422
	$\geq 180\,\mathrm{cm}$	9	0.7167 ± 0.0421
Weight	$< 65\,\mathrm{kg}$	9	0.7528 ± 0.0184
	65 to 75 kg	9	0.7512 ± 0.0416
	$\geq 75\,\mathrm{kg}$	9	0.7164 ± 0.0418
Age	$< 27\,\mathrm{y}$	7	0.7489 ± 0.0228
	27 to 29 y	10	0.7318 ± 0.0488
	$\geq 29\,\mathrm{y}$	10	0.7424 ± 0.0363

Fig. 11.   Classification accuracy versus the BMI for all 27 subjects as well as the predictions by the linear regressor. Small variations in the predictions are due to differences in the other independent variables among the subjects.

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 were 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 were instructed using short movies that the subjects were asked to imitate. This makes the protocol extremely simple, stress- and fatigue-free for the subjects.

Linear regression analysis of signal variability among movement repetitions highlights that 63% of the subjects decrease the range of movement significantly ($p < 0.05$) while repeating a movement and that 22% of the subjects reduce their muscular activity significantly ($p < 0.05$). This result can be related to neuromuscular adaptation to the movement and it needs to be studied in order to evaluate its effect on movement classification.

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 related work, we found that the nonlinear 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 representations. This result also implies that relatively simple features as MAV can perform just as well as more advanced mDWT or STFT features, provided that they are combined with an appropriate classifier.

The distribution of classification accuracy over both subjects and movements demonstrates that there are no outliers in case of subjects, while there is a single outlier with very high performance in the case of movements. This outlier corresponds to the rest posture. Aided by the fact that it accounts for more than 60% of the analyzed samples, the rest contributes to obfuscate the average accuracy of the other movements. A quantitative estimate of the influence of rest on the average classification accuracy is given by the 25% difference between the mean of the Gaussian fit computed for the subject distribution [Fig. 10(a)] and the mean of the movement distribution [Fig. 10(b)]. The resting position is often included as a separate movement in literature, so this difference can affect other results reported in the literature as well. The Ninapro database gives the opportunity to research groups to compare their classification results on the same database and we strongly recommend the users do describe the analyzed movement setting (rest included) in detail in order to make the results comparable.

Anatomical characteristics (including e.g., fat, limb muscular mass, hairiness, and BMI), characteristics of the hand

movement patterns, fatigue, physical parameters and detection-system parameters contribute to determine the sEMG signal [34], [35], while data processing factors (e.g., filtering, windowing, window timepoint, and feature selection) determine how the characteristics of the signal are evaluated during the classification process. All of these factors can potentially affect the classification accuracy. However, in this paper we show how some of them have effects on the data classification. First, our 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 setting are best directed towards these ambiguous phases of the movements. Second, 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 analysis suggests that this degradation also affects movement classification accuracy. In contrast, the accuracy was not found to depend significantly on subject gender, age, or laterality.

The analysis and the results described in this paper have the goal to be a benchmark for the scientific community regarding the results on the NINAPRO database and an example of how it can be used. All the research groups in the field now have the opportunity to work on this resource and they can exploit the possibility to compare their results on the exactly same basis. Therefore, the NINAPRO database gives the scientific community the opportunity to converge to a common position on hand movement recognition by surface electromyography sEMG. We sincerely hope that this opportunity will be exploited and that it will contribute to improve the capabilities of future hand prostheses.

A. Future Work

1) Growth of the Database: In this paper we describe the first available NINAPRO database, which includes 27 intact subjects. New data sets are currently being acquired in order to increase the number of subjects and in order to include also hand amputated subjects. Moreover, we encourage the addition of new datasets by other research groups using the same protocol in order to increase the overall number of subjects: the acquisition protocol can be reproduced easily and we will evaluate any data submission before publishing it on the Ninapro website[7].

2) Improvements to the Setup: The classification results in Section IV-B indicated that more advanced feature representations such as mDWT did not yield improvements over the simpler MAV or WL features. One possible explanation is that the rectification step on 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

---

[7]http://ninapro.hevs.ch/

text 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 three-axis accelerometer. The latter 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: 1) sEMG can be naturally associated with *graded forces* as well as with movements and postures (see, e.g., [2], [43]); 2) the use of regression rather than classification could 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 [44]. Adaptation and calibration of the new devices is already done. We plan to include data obtained using it in the next database release.

3) Further Considerations: The experience obtained with the acquisition protocol and setup using intact subjects will prove useful in the next phase of acquisitions from amputated persons. Movement 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.

REFERENCES

[1] F. R. Finley and R. W. Wirta, "Myocoder studies of multiple myopotential response," *Arch. Phys. Med. Rehabil.*, vol. 48, no. 11, pp. 598–601, 1967.

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

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

[4] S. Micera, J. Carpaneto, and S. Raspopovic, "Control of hand prostheses using peripheral information," *IEEE Rev. Biomed. Eng.*, vol. 3, pp. 48–68, Oct. 2010.

[5] M. Everingham, L. v. Gool, C. K. I. Williams, J. Winn, and A. Zisserman, The PASCAL visual object classes challenge 2010 (VOC2010) results 2010 [Online]. Available: http://www.pascal-network.org/challenges/VOC/voc2010/workshop

[6] G. Griffin, A. Holub, and P. Perona, Caltech-256 object category dataset California Inst. Technol., Tech. Rep. 7694, 2007 [Online]. Available: http://authors.library.caltech.edu/7694

[7] J. Xiao, J. Hays, K. Ehinger, A. Oliva, and A. Torralba, "Sun database: Large-scale scene recognition from abbey to zoo," in *Proc. IEEE Comput. Vis. Pattern Recognit. Conf*, 2010, pp. 3485–3492.

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

[9] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of RGB-D SLAM systems," in *Proc. Int. Conf. Intell. Robot Syst.*, Oct. 2012, pp. 573–580.

[10] H. Müller, A. Rosset, J.-P. Vallée, F. Terrier, and A. Geissbuhler, "A reference data set for the evaluation of medical image retrieval systems," *Comput. Med. Imag. Graph.*, vol. 28, no. 6, pp. 295–305, 2004.

[11] M. Atzori, A. Gijsberts, S. Heynen, A.-G. Mittaz-Hager, O. Deriaz, P. v. d. Smagt, C. Castellini, B. Caputo, and H. Müller, "Building the NINAPRO database: A resource for the biorobotics community," in *Proc. IEEE Int. Conf. Biomed. Robot. Biomechatron.*, 2012, pp. 1258–1265.

[12] L. J. Hargrove, K. Englehart, and B. Hudgins, "A comparison of surface and intramuscular myoelectric signal classification," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 5, pp. 847–853, May 2007.

[13] M. A. Oskoei and H. Hu, "Support vector machine-based classification scheme for myoelectric control applied to upper limb," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 8, pp. 1956–1965, Aug. 2008.

[14] E. Scheme, K. Englehart, and B. Hudgins, "Selective classification for improved robustness of myoelectric control under nonideal conditions," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 6, pp. 1698–1705, Jun. 2011.

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

[16] 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 Trans. Biomed. Eng.*, vol. 56, no. 5, pp. 1427–1434, May 2009.

[17] 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 Trans. Soc. Instrum. Control Eng.*, vol. 29, no. 10, pp. 1213–1220, Oct. 1993.

[18] O. Fukuda, T. Tsuji, M. Kaneko, and A. Otsuka, "A human-assisting manipulator teleoperated by EMG signals and arm motions," *IEEE Trans. Robot. Automat.*, vol. 19, no. 2, pp. 210–222, Apr. 2003.

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

[20] M. Atzori, C. Castellini, and H. Müller, "Spatial registration of hand muscle electromyography signals," presented at the 7th Int. Workshop Biosignal Interpretation, Como, Jul. 2012.

[21] 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)," *Am. J. Ind. Med.*, vol. 29, no. 6, pp. 602–608, 1996.

[22] J. L. G. Nielsen, S. Holmgaard, N. Jiang, K. B. Englehart, D. Farina, and P. A. Parker, "Simultaneous and proportional force estimation for multifunction myoelectric prostheses using mirrored bilateral training," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 3, pp. 681–688, Mar. 2011.

[23] 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," *J. Hand Surg.*, vol. 30, no. 4, pp. 780–789, Jul. 2005.

[24] 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 Trans. Biomed. Eng.*, vol. 55, no. 3, pp. 2198–2211, Mar. 2008.

[25] T. Feix *et al.*, "A comprehensive grasp taxonomy," in *IROS Workshop Understand. Human Hand Adv. Robot. Manipulat.*, 2009, pp. 58–59.

[26] M. R. Cutkosky, "On grasp choice, grasp models, and the design of hands for manufacturing tasks," *IEEE Trans. Robot. Automat.*, vol. 5, no. 3, pp. 269–279, Jun. 1989.

[27] N. Kamakura, M. Matsuo, H. Ishii, F. Mitsuboshi, and Y. Miura, "Patterns of static prehension in normal hands," *Am. J. Occupat. Ther.*, vol. 34, no. 7, pp. 437–445, 1980.

[28] S. J. Edwards, D. J. Buckland, and J. D. McCoy-Powlen, *Developmental and Functional Hand Grasps*. Thorofare, NJ: Slack, 2002.

[29] R. Kato, H. Yokoi, and T. Arai, "Competitive learning method for robust EMG-motion classifier," in *Proc. Intell. Auton. Syst.*, 2006, pp. 946–953.

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

[31] G. Staude, "Objective motor response onset detection in surface myoelectric signals," *Med. Eng. Phys.*, vol. 21, no. 6-7, pp. 449–467, 1999.

[32] J. D. Hamilton, *Time Series Analysis*. Princeton, NJ: Princeton Univ. Press, 1994.

[33] L. Smith, L. Hargrove, B. Lock, and T. Kuiken, "Determining the optimal window length for pattern recognition-based myoelectric control: Balancing the competing effects of classification error and controller delay," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 19, no. 2, pp. 186–192, Apr. 2011.

[34] T. A. Kuiken, M. M. Lowery, and N. S. Stoykov, "The effect of subcutaneous fat on myoelectric signal amplitude and cross-talk," *Prosthet. Orthot. Int.*, vol. 27, no. 1, pp. 48–54, 2003.

[35] D. Farina, C. Cescon, and R. Merletti, "Influence of anatomical, physical, and detection-system parameters on surface EMG," *Biol. Cybern.*, vol. 86, no. 6, pp. 445–456, 2002.

[36] M. Zecca, S. Micera, M. C. Carrozza, and P. Dario, "Control of multifunctional prosthetic hands by processing the electromyographic signal," *Crit. Rev. Biomed. Eng.*, vol. 30, no. 4-6, pp. 459–485, 2002.

[37] M. Lucas, A. Gaufriau, S. Pascual, C. Doncarli, and D. Farina, "Multichannel surface EMG classification using support vector machines and signal-based wavelet optimization," *Biomed. Signal Process. Control*, vol. 3, no. 2, pp. 169–174, 2008.

[38] I. Kuzborskij, A. Gijsberts, and B. Caputo, "On the challenge of classifying 52 hand movements from surface electromyography," in *Proc. 34th Annu. IEEE Eng. Med. Biol. Soc. Conf*, 2012, pp. 4931–4937.

[39] K. Englehart and B. Hudgins, "A robust, real-time control scheme for multifunction myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 7, pp. 848–854, Jul. 2003.

[40] K. Englehart, B. Hudgins, P. A. Parker, and M. Stevenson, "Classification of the myoelectric signal using time-frequency based representations," *Med. Eng. Phys.*, vol. 21, no. 6-7, pp. 431–438, 1999.

[41] H. Bouwsema, C. K. v. d. Sluis, and R. M. Bongers, "Learning to control opening and closing a myoelectric hand," *Arch. Phys. Med. Rehabil.*, vol. 91, no. 9, pp. 1442–1446, 2010.

[42] E. Criswell, *Cram's introduction to surface electromyography*. Sudbury, MA: Jones Bartlett Learning, 2010.

[43] C. Castellini and R. Kõiva, "Using surface electromyography to predict single finger forces," in *Proc. IEEE Int. Conf. Biomed. Robot. Biomechatron.*, 2012, pp. 1266–1272.

[44] R. Kõiva, B. Hilsenbeck, and C. Castellini, "FFLS: An accurate linear device for measuring synergistic finger contractions," in *Proc. 34th Annu. Conf. IEEE Eng. Med. Biol. Soc.*, 2012, pp. 531–534.