Article

Regulating Grip Forces through EMG-Controlled Protheses for Transradial Amputees

Irati Rasines [1,2,*], Miguel Prada [1], Viacheslav Bobrov [3], Dhruv Agrawal [3], Leire Martinez [1], Pedro Iriondo [2], Anthony Remazeilles [1] and Joseph McIntyre [1,4]

1. TECNALIA, Basque Research and Technology Alliance (BRTA), Arteaga Auzoa 25, 48160 Derio, Spain; miguel.prada@tecnalia.com (M.P.); leire.martinez@tecnalia.com (L.M.); anthony.remazeilles@tecnalia.com (A.R.); joseph.mcintyre@tecnalia.com (J.M.)
2. Department of Automatic Control and Systems Engineering, Faculty of Engineering in Bilbao, University of the Basque Country (UPV/EHU), Plaza Ingeniero Torres Quevedo 1, 48013 Bilbao, Spain; pedro.iriondo@ehu.eus
3. Aether Biomedical, Krolowej Jadwigi 43, 61-871 Poznan, Poland; slava@aetherbiomedical.com (V.B.); dhruv@aetherbiomedical.com (D.A.)
4. IKERBASQUE, Basque Foundation for Science, 48013 Bilbao, Spain
* Correspondence: irati.rasines@tecnalia.com or irasines001@ikasle.ehu.es

Abstract: This study aims to evaluate different combinations of features and algorithms to be used in the control of a prosthetic hand wherein both the configuration of the fingers and the gripping forces can be controlled. This requires identifying machine learning algorithms and feature sets to detect both intended force variation and hand gestures in EMG signals recorded from upper-limb amputees. However, despite the decades of research into pattern recognition techniques, each new problem requires researchers to find a suitable classification algorithm, as there is no such thing as a universal 'best' solution. Consideration of different techniques and data representation represents a fundamental practice in order to achieve maximally effective results. To this end, we employ a publicly-available database recorded from amputees to evaluate different combinations of features and classifiers. Analysis of data from 9 different individuals shows that both for classic features and for time-dependent power spectrum descriptors (TD-PSD) the proposed logarithmically scaled version of the current window plus previous window achieves the highest classification accuracy. Using linear discriminant analysis (LDA) as a classifier and applying a majority-voting strategy to stabilize the individual window classification, we obtain 88% accuracy with classic features and 89% with TD-PSD features.

Keywords: classification; force level variation; myoelectric control; pattern recognition; robustness; surface electromyogram (sEMG); transradial amputees

1. Introduction

Paraphrasing the famous German philosopher Kant: "The hand is considered as an extension of the human brain to the outside", hence it has vital importance in daily life activities. The human hand is a prehensile organ necessary to carry out working, recreational and communicative activities in our daily lives. Therefore, upper limb amputees experience many different obstacles in their lives.

Upper limb amputations are not only due to accidents. Diseases such as obesity, diabetes, arthritis or vascular problems can lead to amputations, and because these diseases are becoming more and more prevalent, the number of amputees has increased. In numbers, around 30,000–40,000 amputations were performed in the U.S in 2019 and around 27,000 amputations were reported in England between the years 2015 and 2018 [1]. These values demonstrate an increase in demand for arm prosthetics, the market for which was valued at 697 million USD in 2019 and will reach 1408 million USD by the end of 2023, growing at a magnificent CAGR (12.43%) [2].

To overcome these adversities, prosthetic devices can be an effective solution. But as with any other assistive device, natural interaction between the human and the machine is essential. An ideal prosthetic hand should be intuitive to control and should quickly and reliably detect the intended gesture of the user [3]. But even this is not enough. Imagine carrying an egg in one hand and a glass bottle in the other. If we apply the same force on the egg and on the bottle, one of them will break. A reliable prosthetic hand must not only recognize intended hand posture, it must also successfully decode the user's intended *force* when grasping objects.

An attractive approach to controlling prosthetic devices is to command the device through the activation of remaining muscles, but this is a challenging endeavour. As the remaining muscles of the user's forearm differ considerably among users, the generalization of any decoding algorithm based on measured activation in remnant muscles is not trivial. Machine learning techniques can be seen as the tool that will translate and adapt each individual's remaining muscle response into a concrete command [4,5]. Indeed, feeding a time series of raw EMG directly to a classifier can be tempting, considering the Deep Learning algorithms that are currently trending in the literature [6], but these techniques require a large number of inputs to train a model. We therefore take the more traditional route: sEMG *signals* arrive in windows or batches, *features* are extracted from those windows and those features are used to train a conventional *classifier*.

As our goal is to control a self-contained prosthetic device, we have chosen to focus on time-domain features, from amongst the wide variety features that have been proposed [7], because the computational cost is lower than frequency domain features. We also consider that good results in experiments with able-bodied subjects may nevertheless yield poor results when transferred to amputees. Thus, we refer to the work of Campbell and colleagues [8] who found that classic features (mean absolute value, root mean square, zero crossing rate, slope sign changes and waveform length) and time-dependent power spectrum descriptors proposed in [9,10] (TD-PSD) remain functionally coherent when shifting from able-bodied participants to amputees. Thus, in our analysis we prioritize these features over others proposed in [7].

Analogous to the best feature subset identification, there are also many works in the literature comparing classifiers (see e.g., [11–13]), but few present a systematic quantitative comparison of features [14] and only in [9,10] were the EMG signals obtained from amputees. More importantly, detecting intended applied force has received little attention; with most studies relying on subjects with intact limbs. We therefore set out to extend the state-of-the-art by examining what machine-learning methodologies might be used to control grip force.

We have therefore chosen to use the database created by [10] to compare decoding strategies for gesture + force detection, this being, to our knowledge, the only publicly-available database containing labelled data recorded from amputees containing different force contraction levels. Although our study relies heavily on the previously published works [9,10] we build on these studies in three significant ways: First and foremost, although the original authors' main objective was to improve recognition of different hand gestures *despite* variations in the level of muscle activation, ours main goal is to allow the user to *voluntarily modulate* gripping forces applied by the finger through EMG control in a real life scenario. In pursuing that objective, we select a different set of features than those chosen by [9,10], based on the analysis reported in Campbell et al. [8]. Finally, we propose a novel look-back technique that logarithmically scales and concatenates two consecutive windows to improve reliability.

This document is structured as follows: First, we present the database to be analysed. Then, features and filters to be used are described. Finally, we show the results obtained and we conclude with recommendations for applications and further work.

2. Materials and Methods

2.1. EMG Database

For this investigation, we use the database collected by Al-Timemy et al. [10]. They recruited nine trans-radial amputees with unilateral amputation to participate in their study where six movements including different grip and finger movements were recorded. The gestures available in this database are: thumb flexion, index flexion, fine pinch, tripod grip, hook grip and spherical grip (Figure 1).
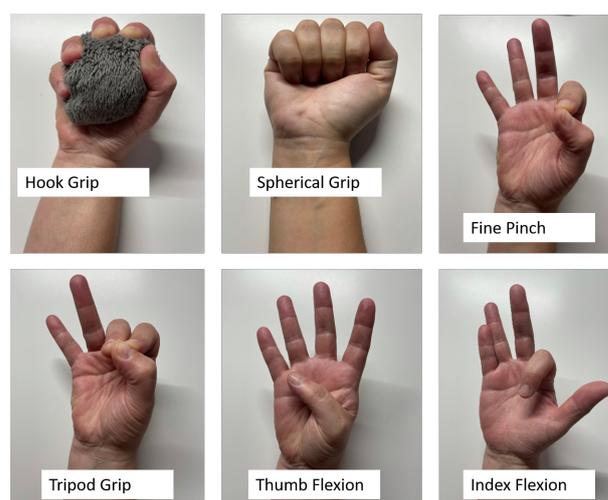
Figure 1. Hand gestures proposed by Al-Timemy et al. in [10].

Each of these gestures was performed at 3 different force levels by the users: low, medium and high.

2.2. EMG Features

We use two different feature groups, which we call *classic* and *TD-PSD* features.

- **Classic Features:** We extracted the following classic state-of-the-art features from the filtered data: Mean Absolute Value (MAV), Root Mean Square (RMS), Zero Crossing Rate(ZC),Slope Sign Changes (SSC) and Waveform Length (WL). We refer the user to [15] for a more extended description of these well known features.
- **Time-dependent power spectrum descriptors:** TD-PSD features are time domain power spectral descriptors proposed by Khushaba et al. [9] and Al-Timemy et al. [10] which include: Moments of the waveform and its first and second derivatives m_0, m_2 and m_4, Power Distribution (S), Irregularity Factor (irr), Waveform Length Ratio (WLR). A single feature vector was constructed composed of the log of each of the aforementioned features.

A single feature vector was constructed for each of these feature groups and used as input to a classifier (described below) to identify the user's intent.

2.3. Windowing and Look-Back Filtering

The selected features were extracted from small windows (150 ms) to simulate the real life scenario where signals arrive in batches. We windowed each trial of the data with overlaps of 50 ms. Instead of randomly associating each individual time window to either the training or test sets, however, we assigned all windows from a given trial to one or the other. This allows for specific processing of consecutive windows as would be the case in real-life scenarios where the user must first train and then use the device. A novelty of this investigation is that not only are the features extracted from the 'current' window employed; a combination of present and past windows is also considered.

For each feature set (classic and TD-PSD) we applied the classification algorithms to the feature vectors computed from a combination of the current and previous windows as follows:

- Present (P): Feature vector extracted from only the current EMG window n:

$$\vec{v}_n = \begin{bmatrix} v_n^1 \\ \vdots \\ v_n^K \end{bmatrix} \tag{1}$$

For the classic features we have $K = 5$ and

$$\vec{v}_n = \begin{bmatrix} \text{MAV} \\ \text{RMS} \\ \text{ZC} \\ \text{SSC} \\ \text{WL} \end{bmatrix} \tag{2}$$

While for the TD-PSD features we have $K = 6$ and

$$\vec{v}_n = \begin{bmatrix} \log m_0 \\ \log(m_0 - m_2) \\ \log(m_0 - m_4) \\ \log S \\ \log F_{\text{irr}} \\ \log R_{\text{WL}} \end{bmatrix} \tag{3}$$

- Log version of Present (PL): Logarithmically scaled version of the previous vector, i.e., $\log(x^2)$:

$$\vec{l}_n = \begin{bmatrix} log[(v_n^1)^2] \\ \vdots \\ log[(v_n^K)^2] \end{bmatrix} \tag{4}$$

- Present-Past (PPa): Feature vector from the current window combined with that of the previous window. Values of the feature vector of the previous window are concatenated with the current one as follows:

$$\vec{b}_n = \begin{bmatrix} \vec{v}_n \\ \vec{v}_{n-1} \end{bmatrix} \tag{5}$$

- Log version of Past and Present (PPaL): Logarithmically scaled $\log(x^2)$ version of the previous vector:

$$\vec{l}_n = \begin{bmatrix} log[(b_n^1)^2] \\ \vdots \\ log[(b_n^{2K})^2] \end{bmatrix} \tag{6}$$

Whereas various techniques reported here have been used previously in other combinations, to our knowledge combining log transformation with look-back (PPaL) is new.

2.3.1. Classifiers

Different machine learning classification algorithms were used in this systematic comparison. They can be classified in 3 big groups:

- Linear Models (4): Logistic Regression (LR), Linear Discriminant Analysis (LDA), Stochastic Gradient Descent(SGD) classifier and Passive Aggressive (PA) classifier.

- Nonlinear Algorithms (5): Decision Trees (DT), Support Vector Machine(with linear kernel) (SVMLinear), Support Vector Machine(with polynomial kernel)(SVMPoly), Support Vector Machine(with rbf kernel)(SVMRbf), Naive Bayes (NB) classifier.
- Ensemble Algorithms (4): Ada Boost Classifier (AB), Bagged Decision Trees (BDT), Random Forest (RF) and Extra Trees (ET) classifier.

All theses algorithms were used with their default hyper-parameter values, except the Ensemble Algorithms for which we set the number of estimators to 100. In this state of the research no hyper-parameter tuning was performed.

In addition to the filtering of the feature vectors as described above, we also implemented a *majority voting* stabilizer that takes the last $N = 20$ outputs of the classifier in the queue and provides the most repeated class as the output.

As stated earlier, each of the six gestures (shown in Figure 1) was performed at three different force levels by the users: low, medium and high. The original authors' main objective was to improve performance of gesture recognition despite force/amplitude variations EMG signals recorded from amputees. Our objective, however, is to allow the user to voluntary modulate gripping forces applied by the fingers of a prosthetic device. We therefore treated the six gestures × three force levels as if they were independent gestures, i.e., as 18 classes. To ease comparison, we mimicked the train/test split used in [10], where 3 trials are used for training each gesture+force state, and the rest are used for testing. As five to eight trials were recorded for each gesture+force condition, this means 54 trials were used for training and from 42 to 99 trials were used for testing, depending on the user. The database contains 1027 total trials (486 for training 541 for testing) and the minimum number of windows a trial contains is 80.

2.3.2. Evaluation Methodology

Many measurements have been proposed in the literature to evaluate the 'goodness' of the results, the most popular being: accuracy rate, f-measure, ROC area under curve and time spent for classification (see [16]). We use accuracy to directly compare our results with other investigations. As the specifics of amputation for each participant is different and the remaining muscular activity can be different, we analysed each user independently, as well as through ensemble averages. We assembled the outcomes of the different combinations of feature vector, classifier and filtering into tabular form and make initial assessments through visual inspection. We then used multi-factor ANOVA to quantitatively compare the efficacy of the most promising solutions.

3. Results

The tables in this section show the results obtained vy averaging the results of the nine users for each independent feature group (classic or TD-PSD) and different filtering methods before (S) and after applying the majority vote (MV) to stabilize the class obtained for each individual window. For visibility purposes, mean values lower than 50% accuracy are omitted. Values shown in color indicate accuracy equal to or above 85%. We further illustrate the results from the most promising algorithms (the same for both feature groups) as bar-plots overlaid with data points from individual users.

3.1. Classic Features

Table 1 summarizes the results obtained for the algorithms tested with classic features and feature combinations and Figure 2 shows each user's result around the mean for the most promising algorithm-feature combination.

- **Linear models:** One can observe in Table 1 the number of blank cells presented in the block of linear models of the table. It is remarkable that the Stochastic Gradient Descent and Passive Aggressive do not achieve more than 50% accuracy for any of the feature combinations as applied to this dataset. The Logistic Regression algorithm achieves greater than 50% accuracy only after majority voting has been applied. Linear

Discriminant Analysis appears to provide the best results out of all classifiers: 78% accuracy for single window analysis; 88% with majority voting.

- **Non-linear models:** Neither Decision Trees, polynomic or RBF versions of Support Vector Machines, nor the Naive Bayes classifier achieve 80% accuracy for the 18 classes. Only the linear version of support vector machine reaches 82% accuracy after majority voting with the PPaL feature combination. Accuracy is lower than with LDA in all conditions.
- **Ensemble models:** The Ada Boost classifier does not provide results better than 50% in any of its configuration. Although the other three proposed classifiers (Bagged Decision Trees, Random Forest, Extra Trees) provide better results, none surpassed 73% even with majority voting.

Individual User Performance

In Figure 2 one can observe that User 5 manifests the poorest results in all the algorithm-feature combinations (its general accuracy is lower than 60% even if the rest of the users cluster around 80%). Additionally, one can observe that user 9 outperforms the rest of the users in terms of accuracy almost achieving perfection with the LDA classifier both for PL and PPaL feature combination with majority voting.

Table 1. Mean of 9 users, Algorithms vs Accuracy of CLASSIC features. 18 gestures without (S) and with majority voting (MV) to stabilize the results.

|	P		PL		PPa		PPaL	
	S	MV	S	MV	S	MV	S	MV
Linear Models								
Logistic Regression		0.58				0.59		0.61
Linear Discriminant Analysis	0.62	0.73	0.74	**0.86**	0.65	0.73	0.78	**0.88**
Stochastic Gradient Descent								
Passive Aggressive								
Non-Linear Models								
Decision Tree		0.63		0.68		0.63		0.68
Support Vector Machine('linear')	0.58	0.67	0.65	0.79	0.60	0.68	0.70	0.82
Support Vector Machine('poly')				0.62			0.50	0.65
Support Vector Machine('rbf')				0.62				0.61
Gaussian Naive Bayes		0.57	0.53	0.62	0.52	0.57	0.56	0.65
Ensemble Models								
Ada Boost								
Bagged Decision Tree	0.58	0.67	0.59	0.73	0.59	0.66	0.62	0.72
Random Forest	0.59	0.68	0.61	0.73	0.61	0.67	0.63	0.72
Extra Trees	0.60	0.67	0.61	0.73	0.61	0.67	0.64	0.72
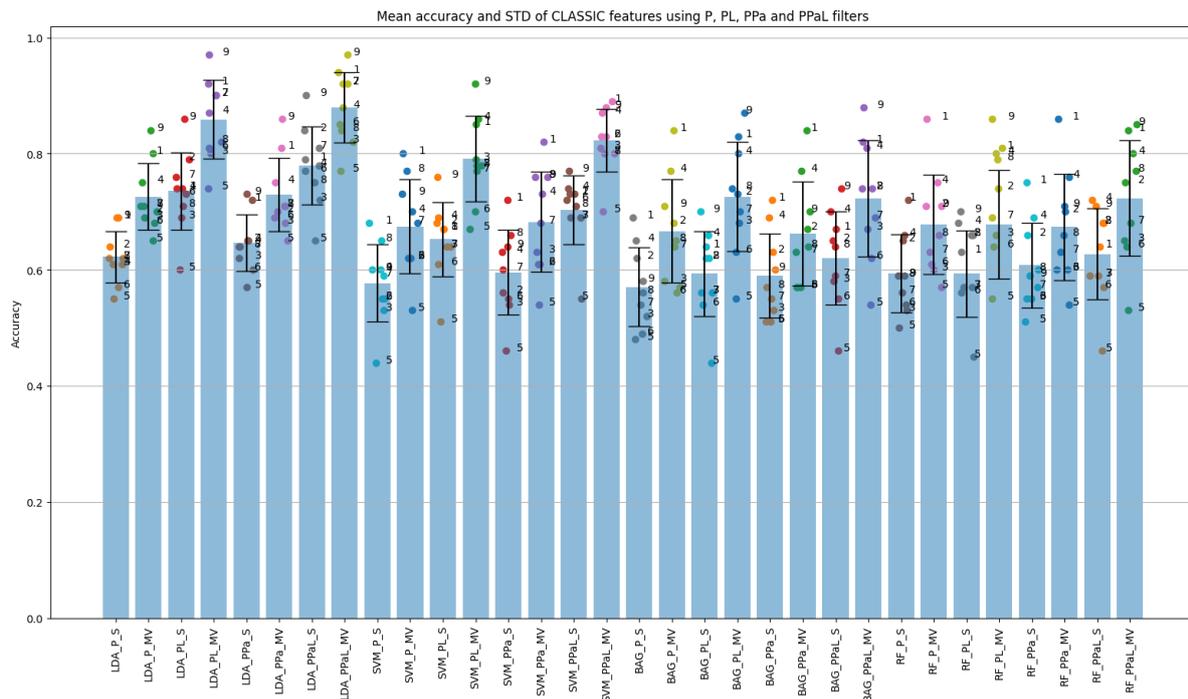
Figure 2. Accuracy value distributions across users for the most promising combinations of Table 1.

3.2. TD-PSD Features

There is a small overall improvement of the obtained accuracies when using TD-PSD features compared to the classic features (i.e., fewer white blocks in Table 2). We describe the results for each algorithm block, as for the classic features above.

- **Linear models:** LDA classifier is once again the algorithm that outperforms the rest of the models of this block. With the TD-PSD features, we can observe easily that the PL and PPaL feature combinations demonstrate the highest accuracy (87% and 89%) after majority voting, 72% and 77% respectively for individual windows.

- **Non-Linear models:** The tendency is similar in the non-linear models when using TD-PSD features compared with classic features. Although in this case we obtain higher general accuracies, only with the SVM with linear kernels is near the auto-imposed 85% percent threshold.

- **Ensemble models:** As in the case of classic features, Bagging Classifiers, Extra trees and Random Forest are the ones providing higher accuracy results. Ada Boost again failed to classify correctly more than half of the instances.

Individual User Performance

Similar tendencies are observed with user 5 and 9 with this feature group (Figure 3). Again user 9 outperforms the rest of the users in terms of accuracy in most of the combinations and user 5 goes below 60% in most of the combinations.

Table 2. Mean of 9 users, Algorithms vs Accuracy of TD-PSD features. 18 gestures wuthout (S) and with majority voting (MV) to stabilize the results.

	P		PL		PPa		PPaL	
	S	**MV**	**S**	**MV**	**S**	**MV**	**S**	**MV**
Linear Models								
Logistic Regression	0.54	0.64	0.55	0.79	0.57	0.65	0.62	0.82
Linear Discriminant Analysis	0.64	0.71	0.72	**0.87**	0.66	0.72	0.77	**0.89**
Stochastic Gradient Descent					0.52			
Passive Aggressive								
Non-Linear Models								
Decision Tree	0.50	0.67	0.54	0.79	0.50	0.66	0.56	0.79
Support Vector Machine('linear')	0.60	0.70	0.54	0.74	0.63	0.71	0.65	0.82
Support Vector Machine('poly')	0.58	0.67	0.49	0.65	0.60	0.66	0.53	0.64
Support Vector Machine('rbf')								
Gaussian Naive Bayes	0.53	0.62	0.61	0.79	0.54	0.61	0.66	0.79
Ensemble Models								
Ada Boost								
Bagged Decision Tree	0.60	0.72	0.66	0.83	0.63	0.72	0.73	**0.85**
Random Forest	0.62	0.71	0.69	0.84	0.64	0.71	0.73	0.84
Extra Trees	0.62	0.71	0.69	**0.85**	0.64	0.71	0.74	0.84
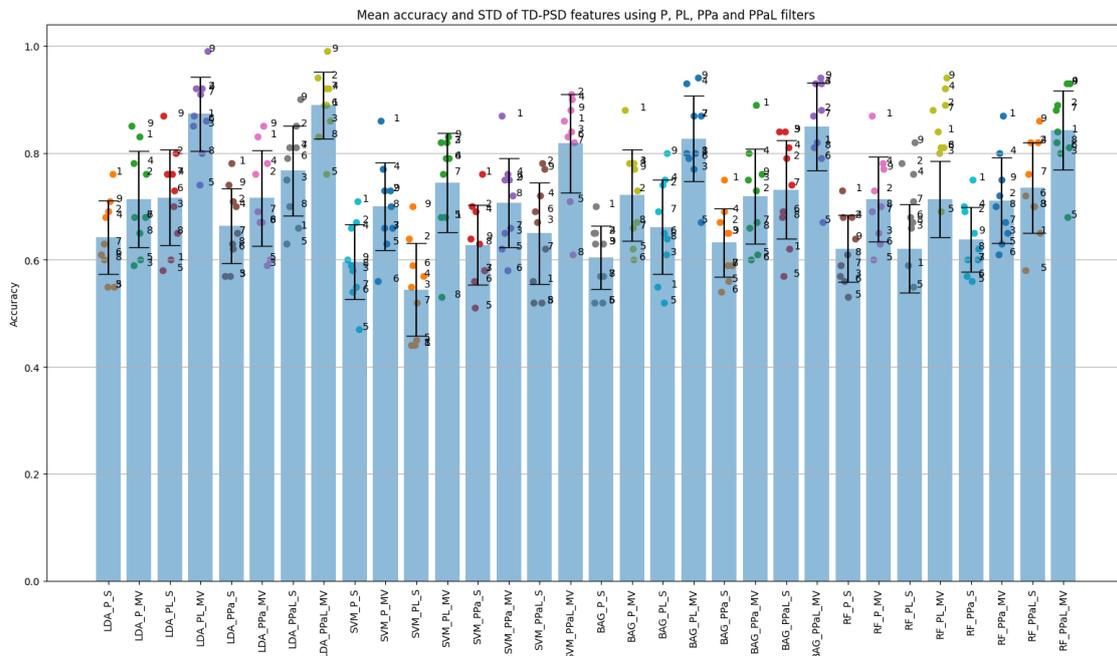
Figure 3. Accuracy value distributions across users for the most promising combinations of Table 2.

3.3. Statistical Comparison

To quantitatively compare methodologies we performed a five-factor ANOVA with repeated measures, with accuracy as the sole dependent variable and with Feature Vector (Classic or TD-PSD), Look-Back Filter Method (P, PPa), Transform (Linear or Log), Majority

Voting (Simple or with MV) and Classifier (LDA, SVM, BAG and RF) as independent factors. We choose to concentrate on these 4 classifiers as they appear to give more promising results than the other options.

Figure 4 shows the relative performance between methods. ANOVA shows a marginally significant main effect of the choice of feature vector ($p < 0.019$) and significant main effects for the other four independent factors ($p < 0.001$), suggesting that (a) the TD-PSD performs better than the Classic feature vector (NB see post-hoc analysis below), (b) LDA gives the best performance amongst the different classifiers, (c) logarithmic transformation of feature components has a significant positive effect on accuracy and (d) both look-back filtering and majority voting generally increase performance.
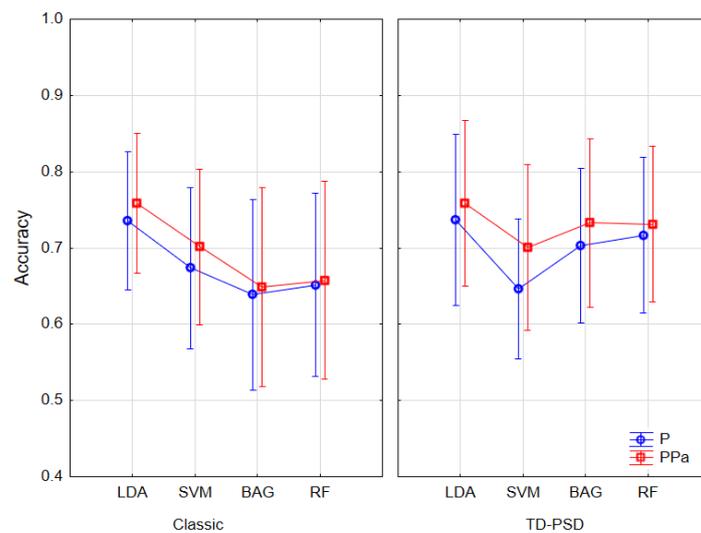

Figure 4. Interaction of Feature Vector, Look-Back Filter Method and Classifier on prediction accuracy.

Applying Tukey's HSD test to identify homogeneous groups, LDA coupled with the log transformed combination of the present and past samples (PPaL) and majority voting stands out as the best combination, with no significant difference between the choice of feature vector (Classic or TD-PSD). Figure 5 showing the results for only the log-tranformed feature elements illustrates this observation. Nevertheless, one can observe in Figure 6 that the accuracy rates appear to vary less across subjects (smaller range of values) for the Classic vs. TD-PSD features.
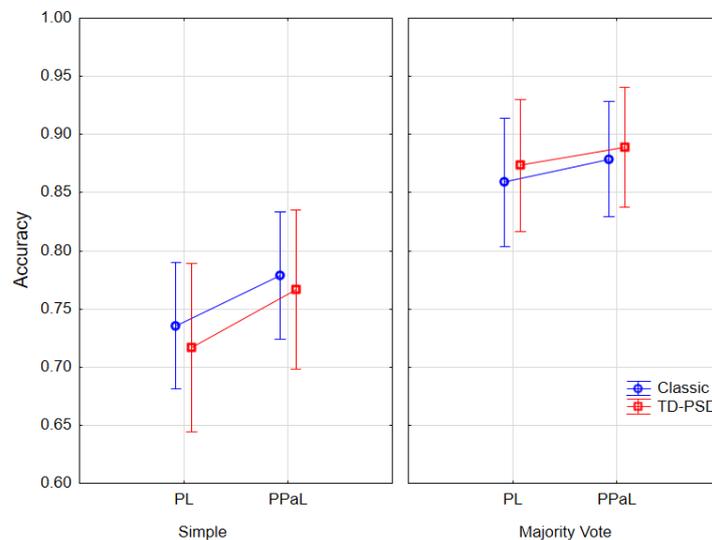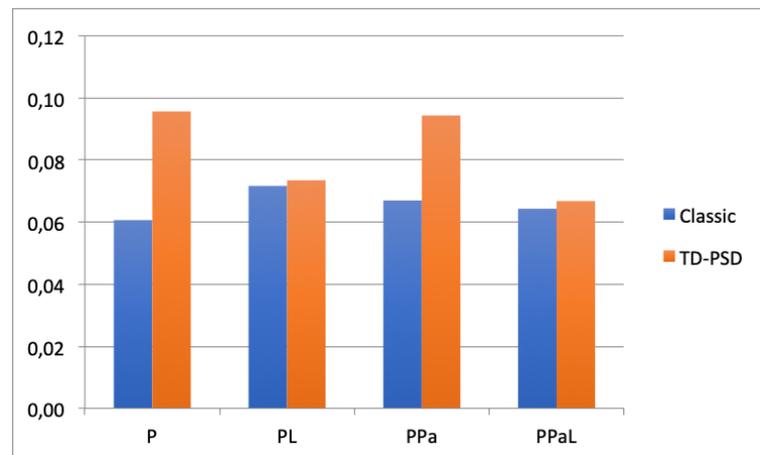

Figure 5. Specific ANOVA analysis of LDA.

Figure 6. Inter-subject variability for different filters. Smaller values indicate greater consistency in performance across individual subjects.

4. Discussion

What novel insights can one draw from our analyses? First, despite showing an advantage over all in the analysis (statistical main effect of Feature Vector), we find no significant advantage of the TD-PSD feature vector over the more classical features when looking specifically at the optimal combination of log transform, look-back and majority voting is applied. This is not entirely surprising. The spectral moments m_0, m_2 and m_4 in the TD-PSD vector are related to the power (RMS), zero crossings (ZC) and slope sign changes (SSC) of the Classic feature vector and both vectors include a measure of waveform length (WV and R_{WL}). The TD-PSD vector includes additional parameters related to Sparseness (S) and Irregularity factor (I_{irr}), but since these two measures are derived from m_0, m_2 and m_4, they do not necessarily bring any new information that can be exploited by the classifier. The TD-PSD features may nevertheless be attractive in a practical case because they rely less on empirical constants (e.g., ϵ thresholds) that could vary from user to user. Note nevertheless that the TD-PSD features include a λ factor that is also empirical (see [9]). In terms of robustness, it would presumably be better if the machine learning algorithm could find the optimal value for this parameter. With a new database in hands, we would propose to perform a grid search for different λ values. This could be done by methodically evaluating the accuracy changes using different λ values and choosing the one that outperforms over users.

Second, our systematic comparison across different classifiers reveals that LDA is reliably better than the three other methods subjected to our statistical analysis (and presumably better than the others as well). One should note, however, that the SVM, BAG and RF methods are not terribly far behind. If they are easier to compute, they could be considered as well.

Third, combining the present and past sample (PPa) and applying majority voting (MV) each individually improved performance, however, there is little improvement gained by combining present and past when majority voting is applied. This makes sense, as both of these methods represent a means to look further back in time. MV is another form of low-pass filtering that will reject spurious classifications when neighbouring values converge to a different categorization. Like low-pass filtering of continuous signals, however, PPa and MV will add latency to the detection of actual state changes, in the MV case depending on the number of values used in the voting. We use $N = 20$ which, for windowing at 150 ms with overlap of 50 ms, leads to a potential delay of up to 1 full second before a state change in the intention of the user is output to the prosthetic device. For the state of the art in which prostheses move slowly, this may be acceptable, but a more detailed analysis of the optimal value for N is also warranted. And as for the PPa look-back filtering, one might also ask if simply widening the time windows or low-pass filtering might also improve accuracy without the added complexity of implementing a voting scheme. Which

of theseto apply will probably also depend on the ease of implementation (PPa is easier to compute than a moving majority vote) and on the desired response latency. Finally, one can consider the consistency across subjects. Consistently lower accuracy rates are obtained for one user than for all the others regardless of the methodology. Possible explanations include low remaining muscle activity, thicker fat layer, incorrect electrode location, etc. Removing this subject (see Figure 6) one might conclude that the Classic feature vector is more robust with respect to inter-user differences, although the data reported here are far to few to state this conclusion with confidence.

4.1. Comparison with Related Studies

The proposed experimental procedure in this investigation could be compared with the experimental Scheme -3 of the creators of the database [10], i.e., training the classifiers with all force levels and testing with single level of force at a time and with the results of the feature fusion proposed in [9]. Al-Timemy et al. [10] report an average error rate when testing with three forces of 17.42% with the LDA classifier, which means a mean accuracy of 82.58%. They tried other classifiers but LDA is the one providing best results in their case, too. In their case, they use the cosine similarity feature vector $f_i = -2a_i b_i / (a_i^2 + b_i^2)$ being a_i the TD-PSD of the i_{th} window and b_i the logarithmically scaled version of a_i as look back filtering instead of the PPaL proposed in this investigation. In [9] graphs suggest an error rate of 18% which translates to 82% accuracy. With our majority voting approach and the logarithmically scaled version of the current and past window (PPaL) combination we achieve 88% accuracy for classic features and 89% for TD-PSD features, outperforming previous attempts despite endeavouring to distinguish between different levels of muscular effort, rather than trying to ignore those variations. This suggests that when the only goal is to identify the gesture, independent of effort, it may nevertheless be more effective to treat different effort/gesture combinations as separate cases in the training process, and then collapse across efforts in the decoding phase, rather than expecting the decoder to ignore variations in overall muscular effort. For our purposes, however, where we would like to offer to the user the ability to control grip forces via changes in muscular effort, these results indicate that this increased level of sophistication in EMG control is indeed achievable.

4.2. Result Transfer to Real World

The present study is a theoretical analysis performed on signals captured in a laboratory environment under favourable circumstances. When transferring the current results to the real life scenario, one must take into account that:

- muscles act under the effects of the gravity due to the weight of the prosthetic device and that interference should be quantified,
- the final user should place the prosthetic on his/her arm without specialized help and this could lead to problems in the EMG readings that do not occur in controlled laboratory conditions,
- the final trained model and the extracted features per window should be light enough to be loaded and extracted on real time in a micro-controllers embedded in a prosthetic hand, and
- the choice of N should be done according to desired latency the capabilities of the EMG capturing system.

Our current investigation points the direction to follow, even if additional work is needed before achieving a robust product.

5. Conclusions

In this paper we present a systematic approach to compare the effects of different choices for feature sets, filtering and classifiers on the accuracy of hand-gesture detection from surface EMG in amputees, where in contrast to most studies, detection of the gripping force intended by the user is also to be decoded. For this objective, at least, Linear Discriminant Analysis proved to be the best classifier compared to a number of linear and

non-linear alternatives, although other classifying methods came in at a close second. We find no clear advantage or disadvantage of using TD-PSD vs. more classical features to characterize the incoming EMG signals so the choice for a given application may depend more on factors not tested here, such as long-term stability and user-to-user variability. Logarithmic scaling and filtering appears to improve accuracy of the classification, as does a scheme of majority voting, albeit at the expense of increased latency to detect state changes. These result show that simultaneous decoding of intended gesture and grip force can be achieved with surface EMG signals from the remnant muscle of an amputated arm. The comparison of feature vectors, classifier methods and filtering presented here provide important insights that may be used to create effective, intuitive and robust control schemes for to meet the needs of transradial amputees.

Author Contributions: Conceptualization, I.R., M.P., V.B., D.A. and J.M.; methodology, I.R., V.B. and J.M.; software, I.R.; formal analysis I.R.; investigation, I.R.; writing—original draft preparation, I.R.; writing—review and editing, I.R., J.M., A.R.; supervision, P.I., L.M.; project administration, L.M. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Data from human users was obtained from a published, on-line database collected under the responsibility of the originating authors [10].

Informed Consent Statement: Data from human users was obtained from a published, on-line database collected under the responsibility of the originating authors [10].

Data Availability Statement: The data used for this investigation can be found at https://www.rami-khushaba.com/electromyogram-emg-repository.html, accessed on 21 October 2021.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Robotic Prosthetic Market Size Description. Available online: https://www.grandviewresearch.com/industry-analysis/robotic-prosthetics-market (accessed on 20 September 2020).
2. Upper Limb Prosthetics Market Size 2023. Available online: https://www.researchandmarkets.com/reports/4618262/prosthetic-arm-market-forecasts-from-2018-to (accessed on 21 October 2021).
3. Englehart, K.; Hudgins, B. A robust, real-time control scheme for multifunction myoelectric control. *IEEE Trans. Biomed. Eng.* **2003**, *50*, 848–854. [CrossRef] [PubMed]
4. Pancholi, S.; Joshi, A.M. Portable EMG data acquisition module for upper limb prosthesis application. *IEEE Sensors J.* **2018**, *18*, 3436–3443. [CrossRef]
5. Cordella, F.; Ciancio, A.L.; Sacchetti, R.; Davalli, A.; Cutti, A.G.; Guglielmelli, E.; Zollo, L. Literature review on needs of upper limb prosthesis users. *Front. Neurosci.* **2016**, *10*, 209. [CrossRef] [PubMed]
6. Jabbari, M.; Khushaba, R.N.; Nazarpour, K. EMG-Based Hand Gesture Classification with Long Short-Term Memory Deep Recurrent Neural Networks. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 20–24 July 2020; IEEE: New York, NY, USA, 2020; pp. 3302–3305.
7. Zecca, M.; Micera, S.; Carrozza, M.C.; Dario, P. Control of multifunctional prosthetic hands by processing the electromyographic signal. *Crit. Rev. Biomed. Eng.* **2002**, *30*, 459–485. [CrossRef] [PubMed]
8. Campbell, E.; Phinyomark, A.; Al-Timemy, A.H.; Khushaba, R.N.; Petri, G.; Scheme, E. Differences in EMG feature space between able-bodied and amputee subjects for myoelectric control. In Proceedings of the 2019 9th International IEEE/EMBS Conference on Neural Engineering (NER), San Francisco, CA, USA, 20–23 March 2019; IEEE: New York, NY, USA, 2019; pp. 33–36.
9. Khushaba, R.N.; Al-Ani, A.; Al-Timemy, A.; Al-Jumaily, A. A fusion of time-domain descriptors for improved myoelectric hand control. In Proceedings of the 2016 IEEE Symposium Series on Computational Intelligence (SSCI), Athens, Greece, 6–9 December 2016; IEEE: New York, NY, USA, 2016; pp. 1–6.
10. Al-Timemy, A.H.; Khushaba, R.N.; Bugmann, G.; Escudero, J. Improving the Performance Against Force Variation of EMG Controlled Multifunctional Upper-Limb Prostheses for Transradial Amputees. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2016**, *24*, 650–661. [CrossRef] [PubMed]
11. Huang, L.C.; Hsu, S.Y.; Lin, E. A comparison of classification methods for predicting Chronic Fatigue Syndrome based on genetic data. *J. Transl. Med.* **2009**, *7*, 81. [CrossRef]

12. Szczuko, P.; Lech, M.; Czyżewski, A. Comparison of classification methods for EEG signals of real and imaginary motion. In *Advances in Feature Selection for Data and Pattern Recognition*; Springer: Cham, Switzerland, 2018; pp. 227–239.

13. Bourouhou, A.; Jilbab, A.; Nacir, C.; Hammouch, A. Comparison of classification methods to detect the Parkinson disease. In Proceedings of the 2016 International Conference on Electrical and Information Technologies (ICEIT), Tangiers, Morocco, 4–7 May 2016; IEEE: New York, NY, USA, 2016; pp. 421–424.

14. Mostafa, S.A.; Mustapha, A.; Mohammed, M.A.; Hamed, R.I.; Arunkumar, N.; Abd Ghani, M.K.; Jaber, M.M.; Khaleefah, S.H. Examining multiple feature evaluation and classification methods for improving the diagnosis of Parkinson's disease. *Cogn. Syst. Res.* **2019**, *54*, 90–99. [CrossRef]

15. Phinyomark, A.; Limsakul, C.; Phukpattaranont, P. A novel feature extraction for robust EMG pattern recognition. *arXiv* **2009**, arXiv:0912.3973.

16. Witten, I.H.; Frank, E. Data mining: Practical machine learning tools and techniques with Java implementations. *Acm Sigmod Rec.* **2002**, *31*, 76–77. [CrossRef]