

RESEARCH ARTICLE

Three-Way Analysis of Spectrospatial Electromyography Data: Classification and Interpretation

Jukka-Pekka Kauppi^{1,2*}, Janne Hahne^{3,4}, Klaus-Robert Müller^{3,5}, Aapo Hyvärinen^{1,6}

1 Dept. of Computer Science/HIIT, University of Helsinki, Helsinki, Finland, **2** Brain Research Unit, O.V. Lounasmaa Laboratory, Aalto University, Espoo, Finland, **3** Dept. of Computer Science, Machine Learning Group, Berlin Institute of Technology, Berlin, Germany, **4** Dept. of Neurorehabilitation Engineering, Universitätsmedizin Göttingen, Göttingen, Germany, **5** Dept. of Brain and Cognitive Engineering, Korea University, Seoul, Korea, **6** Dept. of Dynamic Brain Imaging, Advanced Telecommunication Research Institute International (ATR), Kyoto, Japan

* jukka-pekka.kauppi@helsinki.fi



OPEN ACCESS

Citation: Kauppi J-P, Hahne J, Müller K-R, Hyvärinen A (2015) Three-Way Analysis of Spectrospatial Electromyography Data: Classification and Interpretation. PLoS ONE 10(6): e0127231. doi:10.1371/journal.pone.0127231

Academic Editor: Morniao Xiong, University of Texas School of Public Health, UNITED STATES

Received: December 23, 2014

Accepted: April 12, 2015

Published: June 3, 2015

Copyright: © 2015 Kauppi et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The EMG data sets and the methods presented in this article are available from the following web page: <http://www.cs.helsinki.fi/group/neuroinf/code/spedobox/> and <http://dx.doi.org/10.6084/m9.figshare.1394686>.

Funding: JPK and AH were supported by the Academy of Finland, Centre of Excellence in Inverse Problems (project number 213476), Centre of Excellence in Algorithmic Data Analysis (project number 118653), Computational Sciences (project number 135198), <http://www.aka.fi/en-GB/A/Programmes-and-cooperation/Centres-of-Excellence/>. AH was further supported by the

Abstract

Classifying multivariate electromyography (EMG) data is an important problem in prosthesis control as well as in neurophysiological studies and diagnosis. With modern high-density EMG sensor technology, it is possible to capture the rich spectrospatial structure of the myoelectric activity. We hypothesize that multi-way machine learning methods can efficiently utilize this structure in classification as well as reveal interesting patterns in it. To this end, we investigate the suitability of existing three-way classification methods to EMG-based hand movement classification in spectrospatial domain, as well as extend these methods by sparsification and regularization. We propose to use Fourier-domain independent component analysis as preprocessing to improve classification and interpretability of the results. In high-density EMG experiments on hand movements across 10 subjects, three-way classification yielded higher average performance compared with state-of-the-art classification based on temporal features, suggesting that the three-way analysis approach can efficiently utilize detailed spectrospatial information of high-density EMG. Phase and amplitude patterns of features selected by the classifier in finger-movement data were found to be consistent with known physiology. Thus, our approach can accurately resolve hand and finger movements on the basis of detailed spectrospatial information, and at the same time allows for physiological interpretation of the results.

Introduction

Experimental biomedical data have an intrinsic structure in which data “points” are often matrix valued, i.e. the data should be essentially seen as a multi-way structure. After data collection, temporal measurements are cut to shorter time-windows, also called epochs, and then arranged into a spatio-temporal data array (time × space × epochs). Such three-(rsp. multi-)

Japanese Ministry of Internal Affairs and Communication, project "Novel and innovative R&D making use of brain structures." JH was supported by the Marie Curie IAPP grant "AMYO" (project number 251555), http://cordis.europa.eu/project/rcn/95059_en.html. KRM was supported by the Marie Curie IAPP grant "AMYO" (project number 251555) and the Brain Korea 21 Plus Program as well as the SGER Grant 2014055911 through the National Research Foundation of Korea funded by the Ministry of Education, http://www.nrf.re.kr/nrf_eng.cms/. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

way arrays are common for instance in electromyography (EMG) [1, 2], magnetoencephalography (MEG) [3, 4], electroencephalography (EEG) [5–7] or functional magnetic resonance imaging (fMRI) [8]. Also multimodal data [9] may be arranged in such an array structure. Note that a representation that would squash multi-way data into vectorial shape would clearly ignore the intrinsic spatio-temporal dependency structure induced by the experimental setting and thus miss important information [8, 10, 11].

Applying machine learning methods on multi-way data is often not straight-forward due to these various dependencies, and therefore it is important to study the performance and properties of algorithms in this setting. One task in the analysis of multi-way data sets is to learn an association of the epochs to multiple categories (labels). As alluded above, a three-way classification problem differs from conventional multivariate classification, because each epoch is a matrix (tensor) instead of a vector. Utilizing standard classification methods, such as linear discriminant analysis (LDA) or linear/nonlinear support vector machines (SVMs; see for instance [5]) on vectorized data can yield suboptimal performance due to 1) the very high dimension of the data vector, and 2) the strong dependencies involved in both the signal and noise. Classifying data with special three-way methods which directly use the matrix structure can help to improve performance [10, 12, 13]. For example, bilinear models have been used successfully in the context of brain-computer interfaces [13]. A bilinear classifier is a model that assumes a simplified separable dependency structure typically in space and time.

It is common to enhance the analysis of biomedical data by applying different multivariate techniques, such as principal component analysis (PCA) [14, 15], common spatial pattern (CSP) filtering [1, 16], independent component analysis (ICA) [4, 17–21] and related blind source separation methods [22], to extract signal components with specific spatial characteristics. It is important to note that after such preprocessing, the use of dedicated three-way data analysis methods can still be highly useful, as will be clear below.

In this paper, we concentrate on the classification of three-way data from experimental studies using high-density EMG. High-density electrode arrays are particularly useful for decoding complex hand and/or finger movements, because these arrays can capture information from a high number of signal sources that get activated during corresponding contractions. In contrast to traditional electrode configurations based on only few EMG electrodes, a high-density EMG also allows for estimation of a generative model of source signal activity with much better spatial accuracy [17, 23].

EMG signals are produced by muscle fibers during contraction, and can be used to extract control signals from humans and to give insight to the neuromuscular system [24]. The generation of EMG signals starts by motor commands which are transmitted from the central nervous system to the muscles via action potentials on motoneurons. At the neuromuscular junctions, which are located within the innervation zone typically in the middle of muscle, the action potentials cause depolarizations of muscle cells belonging to one motor unit. These motor unit action potentials (MUAPs) propagate with a speed of approximately 4 m/s towards both ends of the muscle and cause a contraction on the fibers. Due to this propagation, slightly delayed versions of the MUAPs are detected at different channels within the electrode-arrays used in this study. Therefore, a generative model for describing the detected signals should preferably be convolutive [17]. Convolutive models have been previously developed for EEG and MEG data to model phase-lagged activity in different brain areas [19, 25]. In particular, the method called "Fourier-ICA" [19] was developed to model spatially extended sources of rhythmic brain activity in the time-frequency domain with constant phase shifts. Thus, we assume that the Fourier-ICA is capable of detecting the propagation of the MUAPs in high-density EMG arrays due to relatively slow propagation speed of the MUAPs and high spatial resolution of such arrays.

After an amputation, the patient often retains a so called phantom feeling for the missing limb, which he/she can still actuate. For instance, in case of transradial amputation, usually large parts of the forearm muscles that used to control the fingers and wrist remain intact and contract when the phantom limb is actuated. Since the generated EMG signals reflect the muscle-forces, they can be used to extract control information, e.g. to actuate a forearm prosthesis [26]. Most research on this topic has so far focused on relatively simple signal processing schemata where features are extracted from the segmented signals of typically less than ten bipolar electrodes and a subsequent classifier learns the type of movement. Such a procedure is mainly used to extract rough movements such as hand gestures and wrist movements. Modern high-density EMG gives high-dimensional data which has potentially much more information, but extracting such new information may necessitate more sophisticated methods.

Our goal in this contribution is to investigate the suitability of three-way methods for EMG signal classification. We improve and adapt existing methods to make them better fit high-density EMG and its rich spectro-spatial structure. We aim at building a high-accuracy classifier that, importantly, also provides meaningful insight into high-dimensional, three-way EMG data and thus enables a simple interpretation of the physiology behind the classification process. In particular, we first apply the Fourier-ICA algorithm to estimate independent components from the multiarray EMG measurements. Then, we combine a sparsity penalty with a bilinear classifier to improve classification and interpretability of the method. We furthermore develop a computationally efficient and statistically stable extension of linear discriminant analysis to three-way data. In sum, these advances contribute to a significant progress over the existing state-of-the-art in EMG analysis.

Materials and Methods

EMG data

We analyzed data sets from two different EMG movement experiments. The experiments were in accordance with the declaration of Helsinki and were approved by the local ethics commission (Ethikkommission der Charité Berlin, approval number: EA4/085/11). A written informed consent, which was approved by the ethics commission, was obtained from each participant.

Hand-movement data. The first data set was collected previously [1] in an experiment where ten able-bodied subjects repeatedly performed six different hand contractions (hand open, hand close, pronation, supination, wrist flexion, and wrist extension) in five different arm positions. Each contraction lasted about 3 s and was performed with approximately 40% force. Monopolar EMG signals from 96 channels were recorded from the forearms of the subjects using a multiarray EMG grid. For more details of the data and experiment, please see [1]. Our goal was to classify the data set into the six aforementioned movement categories (different arm positions were combined together). We used this data set to quantitatively evaluate the performance of the classifiers.

Finger-movement data. The second data set was collected particularly for this study to demonstrate that our methods are capable of exploiting propagating signal-components that are caused by the underlying composition of the EMG from individual MUAPs. We chose the investigation of individual finger-movements that involve relatively small and adjacent but also superficial muscles located in the forearm. This is not only a challenging classification task which is highly relevant for advanced hand-prostheses, but also an interesting example to demonstrate the interpretability of the generated patterns. We wanted to collect new and interesting data for the purpose of interpretation, because the hand-movement data set has already been examined previously [1].

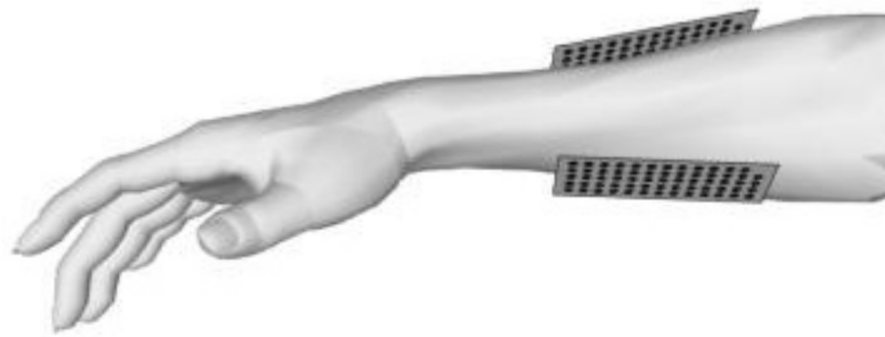


Fig 1. The placement of the high-density EMG sensors in the finger movement experiment.

doi:10.1371/journal.pone.0127231.g001

We recorded 128 monopolar EMG signals from the right hand of one subject, using four “Brain Amp DC” biosignal amplifiers at a sampling frequency of 2500 Hz. This sampling frequency was used to avoid aliasing of the signal (frequency range approx. 0–500 Hz) and to allow for digital filtering. Two Ag/AgCl matrix electrode arrays with an inter-electrode distance of 8 mm (BioOT-ELSCH064R3S) were placed on the skin above the extensors and the flexors of the right (dominant) forearm. The arrays were attached to the skin with foam-pads that contain holes, filled with conductive electrode-paste. They were placed on the forearm (approximately 1/3 from elbow towards the wrist), directly above the muscles that are responsible for the contractions investigated in this study (see Fig 1).

The subject was instructed to perform both isometric flexions (pressings) and extensions (liftings) of the five individual fingers, which resulted in 10 different classes or categories. Further, the subject was instructed to alternate between three different force levels (low/medium/high). Each trial lasted for three seconds. Accurate data labeling was ensured by software that presented the instructions and some feedback to the subject on a computer screen. The subject performed a total of 300 contractions (10 classes, 3 force levels, 10 repetitions). All force levels were combined together in classifier training and testing to mimic variabilities that are present in typical EMG applications.

Preprocessing. Before classification, the data was temporally filtered with a 4th order Butterworth bandpass filter ($f_{low} = 20$ Hz, $f_{high} = 500$ Hz) and a 50 Hz comb-filter to remove movement-artifacts, noise above the EMG band and power-line interferences. The amplitude of the filtered EMG for high contractions was in the range of $185 \mu V_{RMS}$. We performed classification based on 200-ms non-overlapping epochs. For this purpose, we organized hand and finger movement data sets into a three-way spatio-temporal array (time \times channels \times epochs). This lead to 480 epochs per category for the finger movement data and, depending on the subject, 120–210 epochs per category for the hand movement data.

Overview of low-rank bilinear logistic regression

Let us denote a set of epochs (time windows) recorded from multiple sensors in matrix form as:

$$\mathbf{X}(n) = [\mathbf{x}_1(n), \mathbf{x}_2(n), \dots, \mathbf{x}_j(n)]^T \in \mathbb{R}^{D \times E}, \text{ for } n = 1, 2, \dots, N, \quad (1)$$

where the matrix $\mathbf{X}(n)$ denotes the n th epoch, the vector $\mathbf{x}_i(n)$ contains the measurements (time points) from the i -th sensor for the n -th epoch, E is the total number of sensors, D is the number of time-points in one epoch and N is the total number of epochs in the data set. In practice, each epoch may also have been preprocessed by Fourier transformation, replacing the time index by the frequency index.

A simple approach to estimate parameters of the matrix classifier would be to transform matrices to vectors before classification, defining the regression (classifier) function K by:

$$K(\mathbf{G}; \mathbf{X}) = \langle \mathbf{G}, \mathbf{X} \rangle = \text{vec}(\mathbf{G})^T \text{vec}(\mathbf{X}), \tag{2}$$

where $\mathbf{G} \in \mathbb{R}^{D \times E}$ denotes the parameter matrix, $\langle \cdot, \cdot \rangle$ is the elementwise inner product between two matrices, and $\text{vec}(\cdot)$ denotes the vectorization of the matrix columnwise. Because the above function is linear in its parameters, a linear classifier can be trained on these vectors. However, this approach does not utilize the specific structure of the feature matrices and may lead to infeasible parameter estimation problem due to high number of model parameters. To improve the classifier, one can impose a rank constraint on the parameter matrix, such that $\mathbf{G} = \mathbf{W} \mathbf{V}^T$, where $\mathbf{W} = [\mathbf{w}_1 \ \mathbf{w}_2 \ \dots \ \mathbf{w}_R] \in \mathbb{R}^{D \times R}$ and $\mathbf{V} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_R] \in \mathbb{R}^{E \times R}$ are two separable parameter matrices of rank R where R is small. In this case, we have the following *bilinear regression function*:

$$K(\mathbf{W}, \mathbf{V}; \mathbf{X}) = \langle \mathbf{W} \mathbf{V}^T, \mathbf{X} \rangle = \text{Tr}((\mathbf{W} \mathbf{V}^T)^T \mathbf{X}) = \text{Tr}(\mathbf{W}^T \mathbf{X} \mathbf{V}). \tag{3}$$

A binomial logistic regression (LR) can be used to optimize \mathbf{W}, \mathbf{V} for binary classification [13]. The binomial LR model is given by:

$$P(y = 1 | \mathbf{X}) = \frac{1}{1 + \exp(K(\mathbf{W}, \mathbf{V}; \mathbf{X}) + b)} \tag{4}$$

where $P(y = 1 | \mathbf{X})$ is a posterior probability that data matrix \mathbf{X} belongs to category 1, and b is a bias term. Under the assumption that target variables (labels) $y_i \in \{0, 1\}$ are independent and follow a binomial distribution, the normalized log-likelihood of the model is given by:

$$\theta_{\mathbf{w}, \mathbf{v}} = \frac{1}{N} \sum_{i=1}^N (y_i (K(\mathbf{W}, \mathbf{V}; \mathbf{X}(i)) + b) - \log(1 + e^{K(\mathbf{W}, \mathbf{V}; \mathbf{X}(i)) + b})). \tag{5}$$

Sparse low-rank bilinear logistic regression

While the bilinear classifier provides a useful approach to three-way data classification, it does not use the full power of modern learning theory. In particular, sparsity has been shown to improve classification in many forms and many domains. For instance, benefits of utilizing sparsity in continuous arm prosthesis control on the basis of multiarray EMG signals were recently demonstrated [27].

In fact, there is also a deeper reason why a sparsity penalty or similar should be extremely useful for a proper interpretation of bilinear classification. As already pointed out by [13], the solution is highly non-unique, since we have

$$\text{Tr}(\mathbf{W}^T \mathbf{X} \mathbf{V}) = \text{Tr}(\mathbf{U} \mathbf{W}^T \mathbf{X} \mathbf{V} \mathbf{U}^{-1}) \tag{6}$$

for any invertible matrix \mathbf{U} . Thus, the interpretation of the obtained classifier matrices \mathbf{W} and \mathbf{V} is not possible without specifying a principled way of finding the transformation \mathbf{U} . We propose to use the sparsity of the classifier matrices of interest for this purpose.

Thus, we introduce a sparsity penalty on \mathbf{W} and/or \mathbf{V} in the likelihood, and maximize the following objective:

$$\max_{\mathbf{W}, \mathbf{V}} \{ \theta_{\mathbf{W}, \mathbf{V}} + \lambda^w Lp_{\mathbf{W}}(\mathbf{W}) + \lambda^v Lp_{\mathbf{V}}(\mathbf{V}) \}, \tag{7}$$

where $Lp_{\mathbf{W}}(\mathbf{W})$ and $Lp_{\mathbf{V}}(\mathbf{V})$ are regularization terms and at least one of them is the sparsity penalty of the form: $L_1(\mathbf{Z}) = \sum_{ij} |z_{ij}|$. Constants λ^w and λ^v control the amount of regularization and their values are selected through cross-validation. We call the model sparse (low-rank) bilinear classifier.

Optimization of this objective can be difficult due to the L_1 -norm penalty functions which are non-differentiable. However, there exist computationally efficient algorithms for sparse estimation of the linear regression models, suggesting that we can use an “alternating” strategy to maximize the above penalized likelihood. In this strategy, we optimize one parameter matrix at a time while keeping another parameter matrix fixed. For instance, by keeping \mathbf{V} fixed, we can rewrite the regression function as a linear function of \mathbf{W} :

$$K(\mathbf{W}; \mathbf{V}, \mathbf{X}) = \text{Tr}(\mathbf{W}^T \mathbf{X} \mathbf{V}) = \text{Tr}(\mathbf{W}^T \mathbf{Y}) = \langle \mathbf{W}, \mathbf{Y} \rangle \tag{8}$$

where $\mathbf{Y} = \mathbf{X} \mathbf{V}$. Similarly, when \mathbf{W} is fixed, we obtain a linear function of \mathbf{V} . Taking into account the two sparse penalization terms, we actually have the same objectives as in logistic LASSO [28]. If one of the terms is sparse and another one is the traditional L_2 -norm, ridge regression and LASSO problems need to be solved alternately.

To estimate the sparse low-rank bilinear classifier, we can use existing optimization algorithms [29–31] to obtain estimates of the parameter matrices for a sequence of regularization parameters. To find the best regularization parameters, we need to estimate the model for all the possible $m \times n$ regularization parameter pairs $(\lambda_i^v, \lambda_j^w)$ where $i = 1, \dots, m$ and $j = 1, \dots, n$. Putting all these together, our method is as follows:

1. Set rank constraint R and initialize $\mathbf{V}_{\lambda_i^v}$, for $\lambda_1^v, \lambda_2^v, \dots, \lambda_m^v$ to random values.
2. For each $\mathbf{V}_{\lambda_i^v}$ fixed, estimate $\mathbf{W}_{\lambda_j^w}$, for $\lambda_1^w, \lambda_2^w, \dots, \lambda_n^w$ by maximizing (5) with respect to \mathbf{W} only, using the existing methods.
3. For each $\mathbf{W}_{\lambda_j^w}$ fixed, estimate $\mathbf{V}_{\lambda_i^v}$ for $\lambda_1^v, \lambda_2^v, \dots, \lambda_m^v$ likewise.
4. Alternate the above two steps until parameter estimates converge. Then, evaluate the bilinear model for all pairs $(\mathbf{V}_{\lambda_i^v}, \mathbf{W}_{\lambda_j^w})$ using a separate validation data set and pick the best parameters.
5. Compute the test accuracy for an independent data set using the selected model (regularization parameters and classifier parameters) and visualize the classifier.

Thus we obtain a simple method for incorporating a sparsity prior in bilinear classification, utilizing the full power of recently developed sparse classifiers. Generalization for multi-category problems is straightforward using well-known principles, such as one-versus-the-rest (OVR) classification.

Fourier-domain independent component analysis. One of our motivations for using a sparsity enforcing classifier is that when combined with a suitable decomposition (representation) learning method, the classifier allows for the automatic selection of the relevant signal components for physiological investigation. Next, we motivate the ICA decomposition of original raw EMG measurements and propose a method for learning such a decomposition. ICA is a multivariate technique that has been widely used to decompose different physiological signals

into “source” components [32]. The basic assumptions of the ICA model are that 1) the measured signals are a linear mixture of the source components, 2) the sources are independent and non-Gaussian, and 3) the mixing is instantaneous, meaning that a source signal is observed simultaneously in each sensor. Since the components given by ICA are maximally independent, it is likely that the method isolates the information crucial for the classification to a small number of components. Thus, the sparsity-enforcing classifiers are likely to work better when applied on the independent components (ICs) than when applied on the original channels [4].

It has been shown previously that ICA applied in temporal domain of the EMG signal can exploit a rich correlation structure of the multiarray EMG and provide valuable information from these data [17]. However, a major limitation of the basic ICA model in the context of EMG signal analysis is the assumption of instantaneous mixing. In practice, as the MUAPs in muscles are propagating along the muscle-fibers, they are detected in slightly different time on different sensors within the electrode array (in longitudinal direction with respect to the fibers). Therefore it would be better to use a convolutive ICA model, which takes these time-delays into account, to analyze EMG. Previously, the method called Fourier-ICA [19] was developed to model spatially extended sources of rhythmic brain activity in the EEG/MEG signals with constant phase shifts. Because of the high spatial resolution of the high-density EMG and relatively slow propagation speed of the MUAPs, we assume that Fourier-ICA is capable of revealing the propagation of the MUAPs. Fourier-ICA is carried out in the time-frequency domain of the signal and thus allows separate analysis of the magnitudes and phases of the independent components, and in particular it allows for phase shifts (time delays) between the different sensors.

Next, we explain the details of the ICA estimation procedure. At first, we Fourier-transformed the spatio-temporal matrices (time \times channels \times epochs) into spectro-spatial matrices (frequency \times channels \times epochs) in which each column contains the Fourier-coefficients of the sensor in that window. Then, we performed ICA in the Fourier-domain following [19]. For the purpose of ICA, the three-way data was collapsed into an ordinary matrix by concatenating all the epochs and frequencies; the justification for such a procedure is presented in more detail in [19]. Complex-valued ICA was performed on this collapsed matrix (channels \times time \cdot frequency) to estimate ICs. No dimension reduction by PCA was performed. Likewise, the total number of estimated ICs corresponded to the total number of channels of the EMG sensor array.

The resulting ICs were rearranged back into a three-way data but whose dimensions were now: frequency \times ICs \times epochs. Finally, we took the absolute values of all the entries in the three-dimensional data tensor. This is because the tensor contains complex-valued Fourier coefficients, and conventional machine learning algorithms need real-valued input. We also assumed that most of the task-discriminative information is present in the amplitude spectra of the independent components. The phase information was only used in the final interpretation of the results.

Automatic selection of discriminative independent components. We thus combined Fourier-ICA and sparse bilinear classification to extract control information for prosthesis control as well as to automatically find physiologically interesting components from multiarray EMG data. In our study, the term “sparse” reflects the number of zero elements in the classifier parameter matrices \mathbf{V} , \mathbf{W} and not the statistical, non-Gaussian properties or the spatial magnitude patterns of the ICs. The inputs of the classifier are matrices corresponding to epochs (frequency \times ICs), and the parameters to be estimated are the parameter matrix \mathbf{W} for frequency coefficients and the parameter matrix \mathbf{V} for the coefficients of the ICs. After classifier training, the final parameter matrix of the ICs contains non-zero elements only for those ICs that are discriminative in classification. This way, the method automatically returns a discriminative subset of ICs among many redundant and noisy ones. In a binary classification scenario, a

classifier recovers a single subset of task-discriminative ICs. In a multicategory scenario, a separate subset of ICs can be recovered for each category by using e.g. the OVR scheme. Note that the total number of ICs in the estimated classifiers may vary notably between categories and subjects because the total number of non-zero parameters is automatically determined by the algorithm.

Due to its bilinear structure, our classifier makes a simplistic assumption that the estimated ICs share similar spectral characteristics. However, this assumption may not be as restrictive as it seems in the first sight, because it holds only for a (possibly small) subset of discriminative ICs selected by the classifier and not for all the estimated ICs prior to classification. Moreover, by increasing the rank R , other subsets of ICs with different spectral characteristics can be added to the model to improve its flexibility if needed.

Three-way discriminant analysis for dimension reduction

An alternative approach to three-way data classification is to reduce the dimension of the data by computing a small number of components which are particularly relevant for classification, and using them in an ordinary classifier. A number of methods have been proposed recently [12, 33, 34]. Here, we propose a particularly simple and computationally efficient method based on the theory of Fisher's linear discriminant analysis (LDA).

Considering binary classification, denote the difference of the class means by

$$\Delta\mathbf{E} = E\{\mathbf{X}|y = 1\} - E\{\mathbf{X}|y = 0\},$$

where y denotes the class labels. We propose to maximize

$$\max_{\|\mathbf{w}\|=\|\mathbf{v}\|=1} \mathbf{w}^T (\Delta\mathbf{E}) \mathbf{v}$$

which is the basic definition of ordinary LDA with a) the assumption that the class covariances are all equal, and scalar times identity (isotropic) and b) the assumption that the separating vector has matrix form, and rank one as $\mathbf{w}\mathbf{v}^T$. This can be seen as a regularized version of [12], where the covariance matrix is regularized before inversion by adding an identity matrix to it with an infinite weight. Such regularization has the benefit that we completely avoid the computational problems and possible instability related to the inversion of the covariance matrix which is usually done in LDA.

In ordinary LDA, solving this optimization problem we get simply the vector which is the class difference. However, in the three-way case, the solution is given by the dominant singular vectors of $\Delta\mathbf{E}$. Moreover, in the three-way case, it is actually meaningful to compute *many* pairs of singular vectors, while in the ordinary (two-way) case a single vector exhausts all information contained in the class difference. In fact, we can optimize the same objective function again, constraining the new \mathbf{w} and \mathbf{v} to be orthogonal to the previous ones. The solutions are again given by the singular vectors [35].

Thus, we get the following method for binary classification of three-way data:

1. Compute the singular value decomposition (SVD) of $\Delta\mathbf{E}$, the matrix of the differences of the class means.
2. Take the R left (resp. right) singular vectors corresponding to the largest singular values, as the $\mathbf{w}_1, \dots, \mathbf{w}_R$ (resp. $\mathbf{v}_1, \dots, \mathbf{v}_R$).
3. Compute the r -th feature for the n -th data point as $\mathbf{w}_r^T \mathbf{X}(n) \mathbf{v}_r$.

Our approach provides a simple and fundamental way of reducing three-way data into ordinary two-way data. Any ordinary classifier can be used to classify the data based on the features

computed above. We could also incorporate sparsity in this estimation process by using a sparse SVD method [35].

Classifier comparison

We compared the performance of our sparse bilinear (denoted by “Bilinear”) and three-way DA (“ThreeWay DA”) methods against four other classifiers: 1) An ordinary LDA classifier based on Hudgins features and PCA dimension reduction, where the optimal dimension is determined using the CV [36] (denoted by “Hudgins”; this is the state-of-the-art method used in classification of myoelectric signals and serves here as our baseline method); 2) a linear LASSO classifier, for which we vectorized the matrix data prior to classification (denoted by “Linear”); 3) the 2D-LDA dimension reduction method [12], again combined with LASSO-regularized linear logistic regression (denoted by “2D-LDA”); 4) a linear classifier with a trace-norm penalty [10], which imposes a soft low-rank constraint on the parameter matrix (denoted by “TraceNorm”).

As Fourier-ICA is an integral part of our classifier design, we used it to preprocess the original signals prior to classification. The only classifier for which we did not apply Fourier-ICA was “Hudgins”, because it relies on characteristics of temporal EMG signals which would not be detectable from the spectra of the ICs.

We employed five-fold cross-validation to estimate the classification accuracy. For the sparse bilinear classifier, we used the LIGeneral Matlab software with the PSSG optimization [37] for classifier training and performed multicategory classification using the OVR scheme. The features of “ThreeWay DA” were extracted using the OVR scheme and classified using the LASSO-regularized linear logistic regression based on the dual augmented lagrangian algorithm [30]. We fixed the rank $R = 1$ in the estimation of both the three-way DA features and the sparse bilinear model in the comparative study, but we also evaluated the effect of the rank to the classification performance.

Results

Classification accuracy for hand movement data

Fig 2 shows the results of the six tested classifiers for the 6-category hand movement data. Our “Bilinear” (the mean accuracy across subjects was 0.987) classifier yielded the best result together with the “TraceNorm” (the mean accuracy 0.988; according to the paired t -test, there was no statistically significant difference between the accuracies of these methods). The result of “Bilinear” was significantly better when compared with the baseline method “Hudgins” (paired t -test; $p = 0.0019$), which reached the mean accuracy of 0.973. The accuracies of the “ThreeWayDA” (mean accuracy 0.947) and “2D-LDA” (mean accuracy 0.940) were slightly worse than to that of the baseline method. The classifier “Linear” (mean accuracy 0.854) yielded the worst result among the tested methods.

We also evaluated how the choice of rank affects the performance of the classifiers. Fig 3 shows the effect of rank to the classification accuracy for the three relevant methods: “Bilinear”, “ThreeWayDA”, and “2D-LDA”. For the “Bilinear”, the best result was obtained with the rank = 1 and the results degraded slightly with the increase in the rank. For the “ThreeWayDA” and “2D-LDA”, the effect of rank to the classification accuracy was minimal.

Interpretation of finger movement data

The high classification rates above show that three-way classification methods can efficiently utilize spectrospatial information of EMG signals. Next, we demonstrate how the proposed

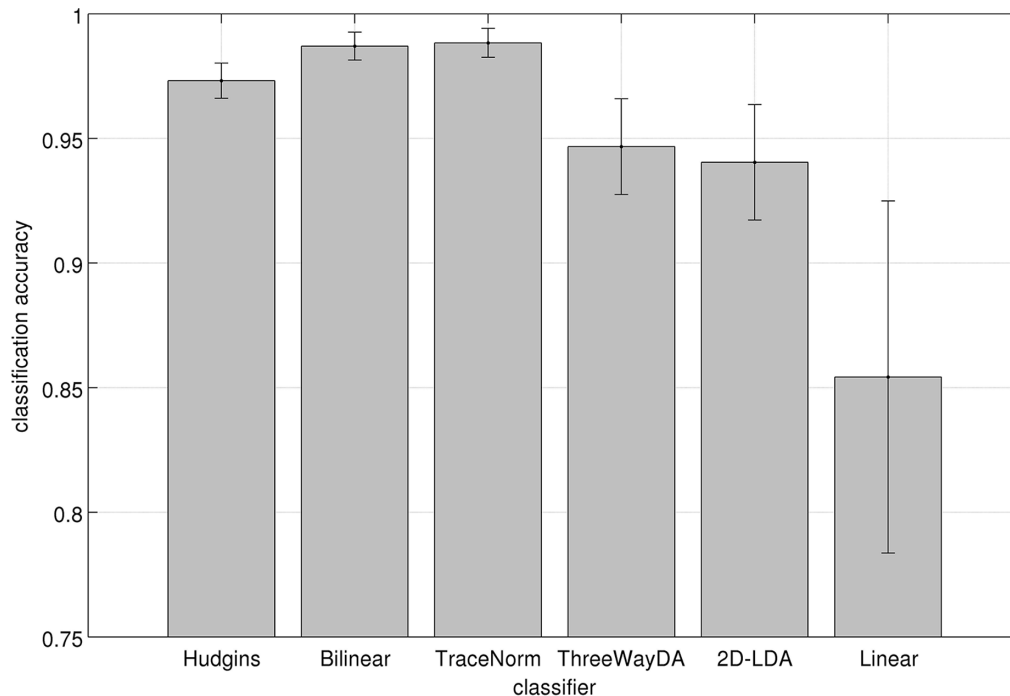


Fig 2. EMG classification results for the 6-category EMG hand movement task. Bars denote the mean (and the standard error of mean) of the classification accuracies across the subjects. The chance level of the classification was 0.167.

doi:10.1371/journal.pone.0127231.g002

sparse bilinear classifier can be used to gain insights into high-dimensional finger movement EMG data (our second data set collected particularly for this purpose). The investigation of EMG activity during finger movements is important as it facilitates the understanding of neurological dysfunctions [38, 39] as well as the optimization of modern myoelectric prosthesis control systems allowing fine-grained movements [40, 41]. The main idea is to investigate the spatial patterns of ICs corresponding to non-zero IC coefficients (given in the columns of \mathbf{V}) of the trained classifier. To ensure that the interpretation of finger movement data is meaningful, we validated that the bilinear classifier was successful also with this data set. For this purpose, we trained the classifier using 2/3 of the data, and evaluated the performance using the remaining 1/3 of the data. The median accuracy across all pairs of finger movement tasks was 0.98, showing that the parameter estimation was highly reliable.

From all the finger movement tasks, we selected one particular task, “lift ring finger” versus “lift little finger”, for closer inspection, by concentrating on the sensor array of 64 electrodes located above the extensors of the forearm. We expected that the analysis of this task is particularly interesting because the classifier must distinguish between activity in relatively small and adjacent muscles located near the sensor array. Fig 4 shows the spatial and phase pattern of the ICs corresponding to non-zero classification coefficients of the sparse bilinear classifier for this task. When interpreting the findings, it is important to keep in mind that while the classifier returns a subset of components which maximally separate between the two categories, it does not mean that the components necessarily have meaningful neuroscientific interpretation, because also artefactual components may improve separability between the categories. For instance here, the amplitudes of ICs #3 and #13 have no clear structure and may be driven by noise. ICs #4, #5, #10 and #20 on the other hand clearly focus on spots that correspond to muscular activity

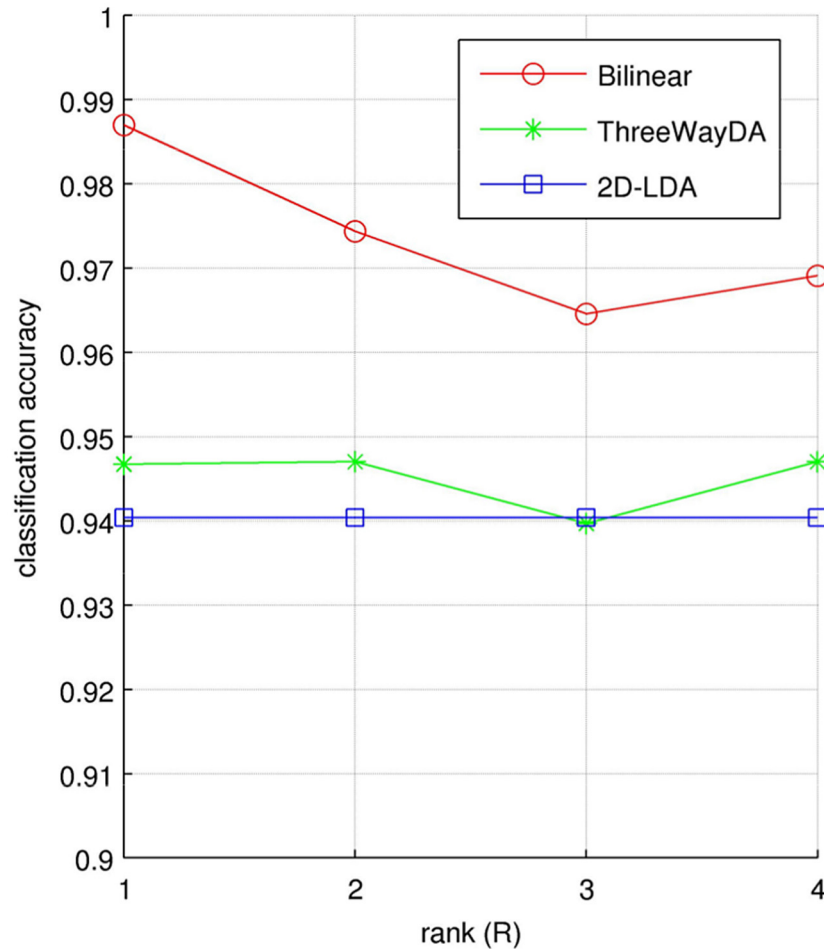


Fig 3. The effect of rank (R) to the classification accuracy for the hand movement data, using the methods “Bilinear”, “ThreeWayDA”, and “2D-LDA”.

doi:10.1371/journal.pone.0127231.g003

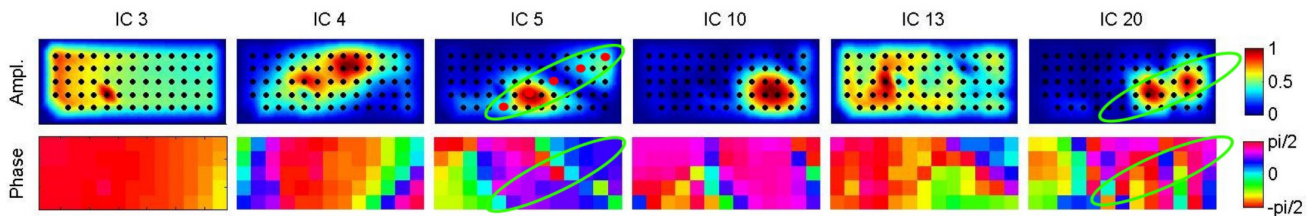


Fig 4. Magnitude and phase patterns of all discriminative ICs found with the sparse bilinear classifier in finger movement data, task “lift ring finger” versus “lift little finger”. The upper row shows the normalized amplitude pattern. Black dots mark the position of the sensors and values in between are obtained by bi-cubic interpolation. The lower row shows the corresponding phase information. Here no interpolation was done because of the periodicity of the phase. Note that phase information is only meaningful in locations where amplitude is non-zero. The sensors showing interesting phase distribution are denoted by the ellipses. The left side of the array is pointing towards the wrist, right side towards the elbow (see Fig 1).

doi:10.1371/journal.pone.0127231.g004

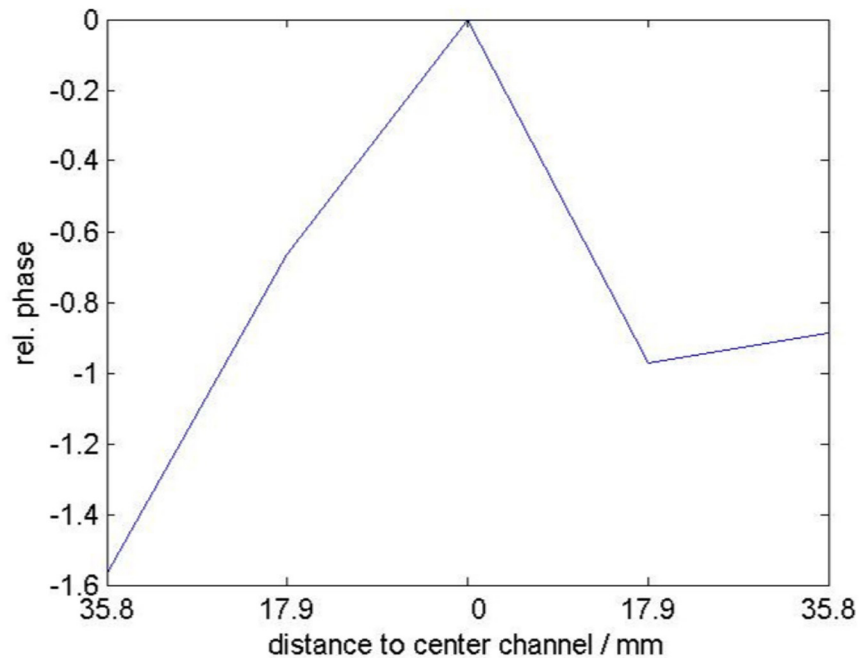


Fig 5. Phase information for five selected channels of IC #5 in Fig 4. The center channel is assumed to be located in the middle of the muscle close to the innervation zone. The x-axis shows the distance to the center channel and the y-axis shows the phase difference relative to the center channel. Assuming one dominant frequency the phase corresponds to a time delay and linear changing phase to propagation with constant speed.

doi:10.1371/journal.pone.0127231.g005

under the sensors. For ICs #5 and #20 one can observe an interesting phase distribution in the sensors denoted by the ellipses: There is a relatively small phase in the center of the assumed muscle position and the phase changes gradually in both directions along the muscle. We hypothesize that this is because of the propagation of the MUAPs from the innervation zone towards the tendons. To investigate this in more detail for IC #5, the phase of some channels on a line parallel to the assumed direction of the muscle fibers (red channels) is shown in Fig 5. Starting from the center, the phase is changing proportionally with the distance to the center electrode with similar slope in both directions. Assuming one dominant frequency on which the ICs mainly react, this corresponds to propagation with constant velocity towards both sides. The fact that the phase does not further change for the last channel on the very right means that the end of the fiber is reached where the MUAPs remain without propagation for a short time before they vanish.

Dimension reduction with three-way DA

Our second tensor method, “ThreeWayDA”, yielded reasonable classification accuracy. We next illustrate how this method can further provide a useful low-dimensional visualization of the EMG data. Figs 6A and 6B show the training and test sets after dimension reduction with the three-way DA for one arbitrarily selected binary task (“press middle finger” vs. “press ring finger”). We used the rank = 2 model to provide a 2-D representation of the original data.

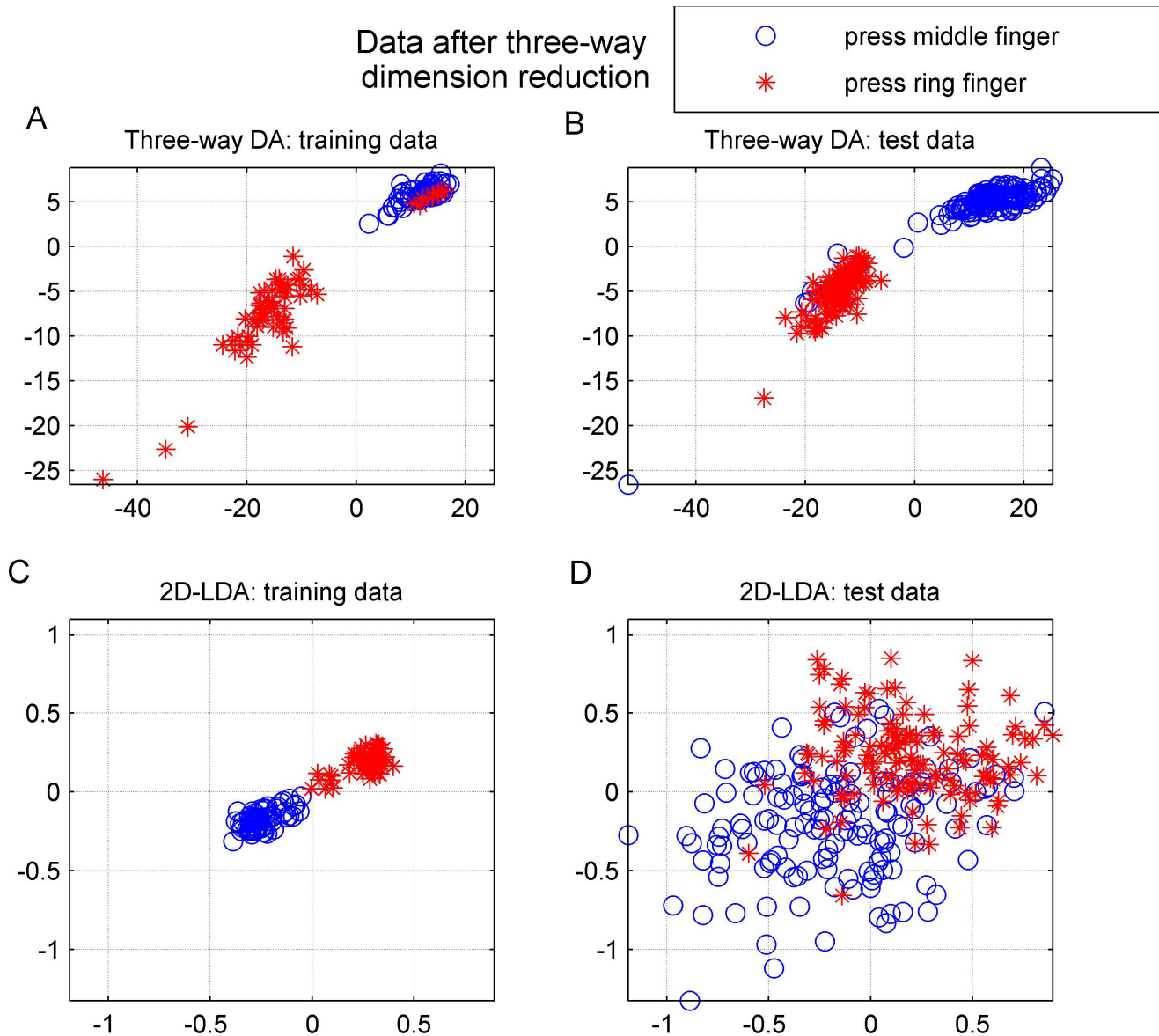


Fig 6. Dimension reduction from data matrix to 2-dimensional data: A) training data for “ThreeWayDA”, B) test data for “ThreeWayDA”, C) training data for “2D-LDA”, D) test data for “2D-LDA”. The horizontal axis corresponds to the first and the vertical axis to the second dimension.

doi:10.1371/journal.pone.0127231.g006

Clearly, the categories were well-separable in a two-dimensional feature space. In fact, the categories were well separable already in 1-D, i.e., the direction of the rank-1 approximation of the mean difference could separate the two classes in this case. The closer analysis of our classifier coefficients showed that the classifier selected a rank-1 solution (i.e., the second classification coefficient was enforced to zero by the LASSO). This example demonstrates that our method can be highly useful in simplification of high-dimensional tensorial EMG data. Because the proposed three-way DA can be seen as a regularized version of 2D-LDA dimension reduction method, it is interesting to compare the features of these methods visually. Figs 6C and 6D

show the corresponding plots for 2D-LDA. This method overfitted the data: it provided excellent separation of the training data but the test data was not as well-separable as in the case of three-way DA.

Discussion

Until now, ordinary two-way signal processing techniques have been used to analyze EMG; none of them employs the rich spectrospatial matrix structure that is present in the multi-way EMG. In this work, we proposed the use of three-way classification methods to multiarray EMG data, as well as contributed algorithmically to three-way classification. In particular, we adapted low-rank bilinear classification to EMG and combined it with sparseness penalties and Fourier-domain independent component analysis to analyze task-discriminative spectrospatial features of the EMG signal. Furthermore, we extended linear discriminant analysis to three-way data for transforming the three-way problem into an ordinary vector classification while at the same time preserving information contained in the multi-way structure. Besides classification, this method allows for visualization of complex high-density EMG data in a low-dimensional feature space.

Three-way classification applied on spectrospatial domain of the EMG signal turned out to be highly successful. Especially, “TraceNorm” and “bilinear” yielded nearly 99% classification accuracy. This result suggests that multiarray EMG consists of fine-grained spectrospatial information that can discriminate between different hand movement tasks. This finding is particularly interesting, because spectral features have not been considered successful in EMG signal classification in a traditional feature extraction setting [42]. A possible explanation for good results in our study is that the three-way classifiers can exploit detailed information from the entire spectrospatial structure of the sensors/components and can simultaneously avoid overfitting due to a low-rank assumption and regularization. In fact, our classification results suggest that *both* sparsity and low-rank constraints are needed to allow accurate classification and meaningful interpretation of EMG data: Clearly, relatively poor result of the classifier based on only the sparsity constraint (“Linear”) shows that sparsity alone is not sufficient for reliable classification. The low-rank constrained classifier (“TraceNorm”) yielded a reliable performance, but it does not allow for simple interpretation of the relevant features because the solution is not sparse. Our sparse and bilinear classifier solves these limitations by combining both the constraints.

Related sparse and bilinear methods have been proposed by [43, 44]. However, our method is clearly different from such previous work. Our primary goal is not to reduce the rank of the matrix through sparsity (i.e., to remove the entire column of the parameter matrices) unlike in [43, 44], but to obtain sparse parameter vectors with only small number of nonzero elements. This is highly useful in the EMG analysis context, where our goal is to associate classification coefficients to the entire spatial patterns of the muscle activity. With too many non-zero coefficients, the interpretation of the final classifiers would be extremely difficult.

Recently, [45] suggested a bilinear modeling approach for hand movement classification from EMG signals. They first extracted features from four EMG electrodes placed around the forearm of the subjects, and then constructed a bilinear model to factorize the measured data into motion- and user-dependent factors. After this, they used the estimated motion-dependent variables and an ordinary SVM classifier to predict the movement category. Our approach is clearly different from this one, because we directly make class predictions on the basis of EMG data by using the bilinear model.

Previously, spatial filtering based on CSP was proposed prior to EMG signal classification [1]. Our new approach differs notably from that one in terms of neurophysiological insights it

provides: The CSP-based approach tells which EMG sensors contain task-discriminative information, whereas the sparse bilinear approach provides a set of EMG activity patterns (characterized by their magnitude and phase) which is most discriminative in a given task. Hence, the two approaches can be used together in future to provide complementary information about the underlying physiological processes of muscles.

Many previous studies have used instantaneous ICA to model EMG signal activity. We estimated ICs in time-frequency domain to better capture characteristics of propagating MUAPs. Under real-world conditions, there are also non-stationaries present in the mixing which are caused e.g. by changing arm positions, movements of the sensor in relation to the muscles, or changing skin-conditions [46]. We did not explicitly model those non-stationarities, but the high classification accuracies regardless of a variation of the arm positions in our hand-movement experiment demonstrated that our methods are relatively robustness against this important source of non-stationarity.

It is also interesting to note that the 2D-LDA, which implements the “full” Fisher discriminant analysis using information in the covariance matrix, provided slightly lower overall accuracy than the simplified, or regularized, version given by our three-way DA. This is presumably because in the 2D-LDA, just as in the general formulation of Fisher LDA for two-way data, the covariance matrix needs to be inverted, or some equivalent tensor operation needs to be performed. Such inversion can be numerically highly unstable, and it can be subject to strong random fluctuations due to finite sample size.

Taken together, the proposed methods can efficiently utilize spectrospatial structure of the high-density EMG data to distinguish between six different hand movements with high accuracy. Furthermore, our investigations with finger lift data suggest that the methods provide classifiers that are interpretable and can thus give novel insights into EMG data. Therefore, our results hold high potential for prosthetics which use EMG to extract control information, as well as for medical diagnosis (cf. [46–48]). Moreover, our methods are not limited to EMG signal analysis, but they can be also applied to other types of biomedical data sets, such as functional brain imaging signals.

Acknowledgments

Correspondence to JPK, JH, KRM and AH. We thank Stefan Haufe for useful discussions.

Author Contributions

Conceived and designed the experiments: JH KRM JPK. Performed the experiments: JH JPK. Analyzed the data: JPK. Contributed reagents/materials/analysis tools: JPK AH KRM JH. Wrote the paper: JPK AH KRM JH. Implemented the new analysis methods: JPK.

References

1. Hahne JM, Graimann B, Müller KR (2012) Spatial filtering for robust myoelectric control. *IEEE Transactions on Biomedical Engineering* 59: 1436–1443. PMID: [22374342](#)
2. Saponas TS, Tan DS, Morris D, Balakrishnan R (2008) Demonstrating the feasibility of using forearm electromyography for muscle-computer interfaces. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, pp. 515–524.
3. van Gerven M, Jensen O (2009) Attention modulations of posterior alpha as a control signal for two-dimensional brain-computer interfaces. *J Neurosci Methods* 179: 78–84. doi: [10.1016/j.jneumeth.2009.01.016](#) PMID: [19428515](#)
4. Kauppi JP, Parkkonen L, Hari R, Hyvärinen A (2013) Decoding magnetoencephalographic rhythmic activity using spectrospatial information. *NeuroImage* 83: 921–936. doi: [10.1016/j.neuroimage.2013.07.026](#) PMID: [23872494](#)

5. Blankertz B, Lemm S, Treder M, Haufe S, Müller KR (2011) Single-trial analysis and classification of ERP components—a tutorial. *NeuroImage* 56: 814–825. doi: [10.1016/j.neuroimage.2010.06.048](https://doi.org/10.1016/j.neuroimage.2010.06.048) PMID: [20600976](https://pubmed.ncbi.nlm.nih.gov/20600976/)
6. Mørup M, Hansen LK, Arnfred SM, Lim LH, Madsen KH (2008) Shift-invariant multilinear decomposition of neuroimaging data. *NeuroImage* 42: 1439–1450. doi: [10.1016/j.neuroimage.2008.05.062](https://doi.org/10.1016/j.neuroimage.2008.05.062) PMID: [18625324](https://pubmed.ncbi.nlm.nih.gov/18625324/)
7. Yu K, Hasan AN, Thakor N, Li X (2014) The analytic bilinear discrimination of single-trial EEG signals in rapid image triage. *PLoS one* 9: e100097. doi: [10.1371/journal.pone.0100097](https://doi.org/10.1371/journal.pone.0100097) PMID: [24933017](https://pubmed.ncbi.nlm.nih.gov/24933017/)
8. Cichocki A, Zdunek R, Phan AH, Amari S (2009) Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation. John Wiley & Sons.
9. Bießmann F, Plis S, Meinecke FC, Eichele T, Müller KR (2011) Analysis of multimodal neuroimaging data. *IEEE Reviews in Biomedical Engineering* 4: 26–58. doi: [10.1109/RBME.2011.2170675](https://doi.org/10.1109/RBME.2011.2170675) PMID: [22273790](https://pubmed.ncbi.nlm.nih.gov/22273790/)
10. Tomioka R, Müller KR (2010) A regularized discriminative framework for EEG analysis with application to brain–computer interface. *NeuroImage* 49: 415–432. doi: [10.1016/j.neuroimage.2009.07.045](https://doi.org/10.1016/j.neuroimage.2009.07.045) PMID: [19646534](https://pubmed.ncbi.nlm.nih.gov/19646534/)
11. Porro-Muñoz D, Duin RP, Talavera I, Orozco-Alzate M (2011) Classification of three-way data by the dissimilarity representation. *Signal Processing* 91: 2520–2529. doi: [10.1016/j.sigpro.2011.05.004](https://doi.org/10.1016/j.sigpro.2011.05.004)
12. Li J, Janardan R, Li Q (2004) Two-dimensional linear discriminant analysis. *Advances in Neural Information Processing Systems* 17: 1569–1576.
13. Dyrholm M, Christoforou C, Parra LC (2007) Bilinear discriminant component analysis. *J Mach Learn Res* 8: 1097–1111.
14. Hargrove LJ, Li G, Englehart KB, Hudgins BS (2009) Principal components analysis preprocessing for improved classification accuracies in pattern-recognition-based myoelectric control. *IEEE Transactions on Biomedical Engineering* 56: 1407–1414. doi: [10.1109/TBME.2008.2008171](https://doi.org/10.1109/TBME.2008.2008171) PMID: [19473932](https://pubmed.ncbi.nlm.nih.gov/19473932/)
15. Casarotto S, Bianchi AM, Cerutti S, Chiarenza GA (2004) Principal component analysis for reduction of ocular artefacts in event-related potentials of normal and dyslexic children. *Clinical neurophysiology* 115: 609–619. doi: [10.1016/j.clinph.2003.10.018](https://doi.org/10.1016/j.clinph.2003.10.018) PMID: [15036057](https://pubmed.ncbi.nlm.nih.gov/15036057/)
16. Blankertz B, Tomioka R, Lemm S, Kawanabe M, Müller KR (2008) Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal Process Mag* 25: 41–56. doi: [10.1109/MSP.2008.4408441](https://doi.org/10.1109/MSP.2008.4408441)
17. Staudenmann D, Daffertshofer A, Kingma I, Stegeman DF, van Dieën JH (2007) Independent component analysis of high-density electromyography in muscle force estimation. *IEEE Transactions on Biomedical Engineering* 54: 751–754. doi: [10.1109/TBME.2006.889202](https://doi.org/10.1109/TBME.2006.889202) PMID: [17405383](https://pubmed.ncbi.nlm.nih.gov/17405383/)
18. Vigário R, Sarela J, Jousmiki V, Hamalainen M, Oja E (2000) Independent component approach to the analysis of EEG and MEG recordings. *IEEE Transactions on Biomedical Engineering* 47: 589–593. doi: [10.1109/10.841330](https://doi.org/10.1109/10.841330) PMID: [10851802](https://pubmed.ncbi.nlm.nih.gov/10851802/)
19. Hyvärinen A, Ramkumar P, Parkkonen L, Hari R (2010) Independent component analysis of short-time Fourier transforms for spontaneous EEG/MEG analysis. *NeuroImage* 49: 257–271. doi: [10.1016/j.neuroimage.2009.08.028](https://doi.org/10.1016/j.neuroimage.2009.08.028) PMID: [19699307](https://pubmed.ncbi.nlm.nih.gov/19699307/)
20. Naik GR, Acharyya A, Nguyen HT (2014) Classification of finger extension and flexion of EMG and Cyberglove data with modified ICA weight matrix. In: 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, pp. 3829–3832.
21. Ren X, Hu X, Wang Z, Yan Z (2006) MUAP extraction and classification based on wavelet transform and ICA for EMG decomposition. *Medical and Biological Engineering and Computing* 44: 371–382. doi: [10.1007/s11517-006-0051-3](https://doi.org/10.1007/s11517-006-0051-3) PMID: [16937179](https://pubmed.ncbi.nlm.nih.gov/16937179/)
22. Jiang N, Farina D (2011) Covariance and time-scale methods for blind separation of delayed sources. *IEEE Transactions on Biomedical Engineering* 58: 550–556. doi: [10.1109/TBME.2010.2084999](https://doi.org/10.1109/TBME.2010.2084999) PMID: [20934944](https://pubmed.ncbi.nlm.nih.gov/20934944/)
23. Blok J, Van Dijk J, Drost G, Zwartz M, Stegeman D (2002) A high-density multichannel surface electromyography system for the characterization of single motor units. *Review of scientific instruments* 73: 1887–1897. doi: [10.1063/1.1455134](https://doi.org/10.1063/1.1455134)
24. Farina D, Jensen W, Akay M (2013) Introduction to neural engineering for motor rehabilitation, volume 40. John Wiley & Sons.
25. Anemüller J, Sejnowski TJ, Makeig S (2003) Complex independent component analysis of frequency-domain electroencephalographic data. *Neural Networks* 16: 1311–1323. doi: [10.1016/j.neunet.2003.08.003](https://doi.org/10.1016/j.neunet.2003.08.003) PMID: [14622887](https://pubmed.ncbi.nlm.nih.gov/14622887/)
26. Oskoei MA, Hu H (2007) Myoelectric control systems—a survey. *Biomedical Signal Processing and Control* 2: 275–294. doi: [10.1016/j.bspc.2007.07.009](https://doi.org/10.1016/j.bspc.2007.07.009)

27. Hwang HJ, Hahne JM, Müller KR (2014) Channel selection for simultaneous and proportional myoelectric prosthesis control of multiple degrees-of-freedom. *Journal of neural engineering* 11: 056008. doi: [10.1088/1741-2560/11/5/056008](https://doi.org/10.1088/1741-2560/11/5/056008) PMID: [25082779](https://pubmed.ncbi.nlm.nih.gov/25082779/)
28. Tibshirani R (1996) Regression shrinkage and selection via the LASSO. *J R Stat Soc: Series B (Stat Meth)* 58: 267–288.
29. Schmidt M (2010) Graphical model structure learning with l1-regularization. Ph.D. thesis, University of British Columbia.
30. Tomioka R, Suzuki T, Sugiyama M (2011) Super-linear convergence of dual augmented lagrangian algorithm for sparsity regularized estimation. *J Mach Learn Res* 12: 1537–1586.
31. Friedman JH, Hastie T, Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 33: 1–22. PMID: [20808728](https://pubmed.ncbi.nlm.nih.gov/20808728/)
32. Hyvärinen A, Karhunen J, Oja E (2001) Independent component analysis. Wiley-Interscience, 1 edition.
33. Tao D, Li X, Wu X, Maybank SJ (2007) General tensor discriminant analysis and gabor features for gait recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29: 1700–1715. doi: [10.1109/TPAMI.2007.1096](https://doi.org/10.1109/TPAMI.2007.1096) PMID: [17699917](https://pubmed.ncbi.nlm.nih.gov/17699917/)
34. Lu H, Plataniotis KN, Venetsanopoulos AN (2009) Uncorrelated multilinear discriminant analysis with regularization and aggregation for tensor object recognition. *IEEE Transactions on Neural Networks* 20: 103–123. doi: [10.1109/TNN.2008.2004625](https://doi.org/10.1109/TNN.2008.2004625) PMID: [19095532](https://pubmed.ncbi.nlm.nih.gov/19095532/)
35. Witten DM, Tibshirani R, Hastie T (2009) A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics* 10: 515–534. doi: [10.1093/biostatistics/kxp008](https://doi.org/10.1093/biostatistics/kxp008) PMID: [19377034](https://pubmed.ncbi.nlm.nih.gov/19377034/)
36. Englehart K, Hudgins B, Parker PA, Stevenson M (1999) Classification of the myoelectric signal using time-frequency based representations. *Medical engineering & physics* 21: 431–438. doi: [10.1016/S1350-4533\(99\)00066-1](https://doi.org/10.1016/S1350-4533(99)00066-1)
37. Schmidt M, Fung G, Rosales R (2007) Fast optimization methods for l1 regularization: A comparative study and two new approaches. In: *Machine Learning: ECML 2007*, Springer. pp. 286–297.
38. Werner RA, Andary M (2011) Electrodiagnostic evaluation of carpal tunnel syndrome. *Muscle & nerve* 44: 597–607. doi: [10.1002/mus.22208](https://doi.org/10.1002/mus.22208)
39. Kimura J (2013) Electrodiagnosis in diseases of nerve and muscle: principles and practice. Oxford University Press.
40. Peleg D, Braiman E, Yom-Tov E, Inbar GF (2002) Classification of finger activation for use in a robotic prosthesis arm. *Transactions on Neural Systems and Rehabilitation Engineering* 10: 290–293. doi: [10.1109/TNSRE.2002.806831](https://doi.org/10.1109/TNSRE.2002.806831) PMID: [12611366](https://pubmed.ncbi.nlm.nih.gov/12611366/)
41. Andrews A, Morin E, McLean L (2009) Optimal electrode configurations for finger movement classification using EMG. In: *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, pp. 2987–2990.
42. Phinyomark A, Phukpattaranont P, Limsakul C (2012) Feature reduction and selection for EMG signal classification. *Expert Systems with Applications* 39: 7420–7431. doi: [10.1016/j.eswa.2012.01.102](https://doi.org/10.1016/j.eswa.2012.01.102)
43. Tan X, Li Y, Liu J, Jiang L (2010) Face liveness detection from a single image with sparse low rank bilinear discriminative model. In: *Computer Vision—ECCV 2010*, Springer. pp. 504–517.
44. Pirsiavash H, Ramanan D, Fowlkes C (2009) Bilinear classifiers for visual recognition. *Advances in Neural Information Processing Systems* 22: 1482–1490.
45. Matsubara T, Morimoto J (2013) Bilinear modeling of EMG signals to extract user-independent features for multiuser myoelectric interface. *IEEE Transactions on Biomedical Engineering* 60: 2205–2213. doi: [10.1109/TBME.2013.2250502](https://doi.org/10.1109/TBME.2013.2250502) PMID: [23475334](https://pubmed.ncbi.nlm.nih.gov/23475334/)
46. Jiang N, Dosen S, Müller KR, Farina D (2012) Myoelectric control of artificial limbs—Is there a need to change focus. *IEEE Signal Processing Magazine* 29: 150–152.
47. Castellini C, van der Smagt P (2009) Surface EMG in advanced hand prosthetics. *Biological Cybernetics* 100: 35–47. doi: [10.1007/s00422-008-0278-1](https://doi.org/10.1007/s00422-008-0278-1) PMID: [19015872](https://pubmed.ncbi.nlm.nih.gov/19015872/)
48. Weiss M (2012) *The Electrodiagnosis of Neuromuscular Disorders, An Issue of Physical Medicine and Rehabilitation Clinics*, volume 24. Elsevier Health Sciences.