Generalized Features for Electrocorticographic BCIs

Pradeep Shenoy*, Kai J. Miller, Jeffrey G. Ojemann, and Rajesh P. N. Rao

***Abstract*—This paper studies classifiability of electrocorticographic signals (ECoG) for use in a human brain–computer interface (BCI). The results show that certain spectral features can be reliably used across several subjects to accurately classify different types of movements. Sparse and nonsparse versions of the support vector machine and regularized linear discriminant analysis linear classifiers are assessed and contrasted for the classification problem. In conjunction with a careful choice of features, the classification process automatically and consistently identifies neurophysiological areas known to be involved in the movements. An average two-class classification accuracy of 95% for real movement and around 80% for imagined movement is shown. The high accuracy and generalizability of these results, obtained with as few as 30 data samples per class, support the use of classification methods for ECoG-based BCIs.**

***Index Terms*—Brain–computer interfaces, classification, electrocorticography, feature selection, neural interfaces.**

I. Introduction

BRAIN–COMPUTER interfaces (BCIs) [1] attempt to provide control of prosthetic or communication devices by direct use of an individual's brain signals. These brain signals can be measured noninvasively (e.g., [2], [3]) in humans using electroencephalography (EEG), and invasively at the level of single neurons and local field potentials in rats and monkeys [4], [5]. Although the invasive BCIs typically outperform EEG-based BCIs, there is significant concern about neural recordings in humans [6] due in part to the invasive nature of the procedure and concerns regarding long-term health risks.

Electrocorticography (ECoG) [7]–[10] has recently gained attention as a recording technique for use in brain–computer interfaces. ECoG involves recording electrical signals from the surface of the human brain, typically in patients being monitored prior to surgery. ECoG is less invasive than neuronal recordings since the brain is not penetrated and has a much higher signal-to-noise ratio (SNR) than EEG, as well as higher spectral and spatial resolution. This higher resolution necessitates re-engineering of the signal processing and classification techniques used in traditional EEG-based BCIs. An obstacle to effectively characterize the information present in ECoG signals is the extreme sparsity of data due to the limited time available for volunteering patients (see Section II).

We study ECoG recordings of 64–104 channels from eight subjects during both overt and imagined movement of the tongue and hand. In contrast to the BCI experiments in the literature, our experiments were limited to only 30 examples of each class. This presents significant challenges for learning, and the danger of overfitting the data. We address these problems by 1) using a single, carefully chosen set of bandpower features for all subjects and 2) using simple linear classifiers, including powerful sparse methods that automatically incorporate the benefits of feature/channel selection. Our results show that overt hand and tongue movements can be distinguished with very high accuracy. Imagined movements are also distinguishable, but with lower accuracy.

We examine the features and channels naively chosen by our classifiers in a posthoc analysis. The classifiers successfully select cortical hand and tongue areas, independently confirming the efficacy of our chosen feature representation. The sparse classification methods choose channels that are more tightly defined over the relevant brain areas and, thus, may perform well in practice as integral components of ECoG-based BCIs. We show that data from overt movements can be used to significantly improve performance on classification of imaginary movements.

A. Related Work

EEG-based BCIs exploit well-characterized electrophysiological phenomena in humans. It is known [11] that signal power in certain spectral bands of individual channels (mu and beta rhythms, 9–13 Hz, 18–24 Hz) varies with motor action or imagery. Further, certain specific channels over the motor cortex in a standardized electrode placement montage have the highest information with reference to motor activity. Several BCIs are designed around this knowledge, and additionally (see, for example, [12]) customize this spatial and spectral feature selection to better fit each individual's data.

There is no broad consensus regarding channel locations or feature representations to use for the classification of ECoG. One significant issue here is that the subject population is undergoing the procedure for medical purposes (see Section II) and, thus, the electrode locations are different in each subject. Another issue is that ECoG is an invasive procedure that is only performed for medical needs. As a result, access to ECoG data and subjects is limited.

Lal *et al.* [7] classify ECoG data from tongue and hand imagined movements. They use autoregressive model coefficients as features for each channel, and recursive channel elimination for selecting the best channels for classification. Graimann *et al.* [8]

Manuscript received November 16, 2006; revised April 18, 2007. This work is based upon work supported by The National Science Foundation under Grants 0622252, 0642848, and 0130705, and a Packard Foundation Fellowship to RPNR. *Asterisk indicates corresponding author.*

*P. Shenoy is with the Department of Computer Science and Engineering, University of Washington, Box 352350, Seattle, WA 98195 USA (e-mail: pshenoy@cs.washington.edu).

K. J. Miller is with the Department of Physics, Computer Science, and Medicine at the University of Washington, Seattle, WA 98195 USA.

J. G. Ojemann is with the Department of Neurosurgery at the University of Washington School of Medicine, Seattle, WA 98195 USA.

R. P. N. Rao is with the Department of Computer Science and Engineering, University of Washington, Seattle, WA 98195 USA (e-mail: kai@cs.washington.edu; jojemann@seattlechildrens.org; rao@cs.washington.edu).

Digital Object Identifier 10.1109/TBME.2007.903528

use wavelet packet analysis and genetic algorithms (GAs) for selecting features related to event-related desynchronization and synchronization (ERD/ERS) that detect single-trial movement of body parts, such as the finger and lip. They have reported the presence of ERD/ERS features in delta (<3.5 Hz), beta (12.5–30 Hz), and gamma (70–90 Hz) bands associated with the onset of a single discrete movement. Leuthardt *et al.* [9] show continuous one-dimensional ECoG-based control by modulation of carefully chosen spectral power features that are selected via a screening task.

II. Experimental Setup

A. Subject Population

Simple motor action and motor imagery tasks were studied in patients with intractable epilepsy with implanted intracranial electrode arrays. These electrodes are implanted in order to localize seizure foci prior to surgical resection of the epileptic focus, and their location is determined independently by clinical criteria. Only patients with some peri-Rolandic coverage were included. Patients underwent craniotomy for electrode placement and were typically studied 4–6 days after placement to allow for recovery from the original surgery.

Each patient typically had an implanted 8×8 electrode platinum electrode array (Ad-Tech, Racine, WI), sometimes accompanied by linear strips, with 1-cm interelectrode distance. The electrodes were embedded in silastic with 2.3-mm diameter exposed (of a 4-mm diameter electrode). The arrays varied in size from a total of 62–104 electrodes across the patients. Fig. 1 shows an example of an implanted ECoG grid and electrode strips. The signal was recorded with Synamps2 (Neuroscan, El Paso, TX) amplifiers at 1000 Hz and bandpass filtered from 0.15 to 200 Hz (well outside the spectral ranges we use for classification). Although 60-Hz signal contamination is ubiquitous in the recordings, our choice of features avoids this contamination. Stimuli are presented and data are collected using the multipurpose BCI2000 software [13].

B. Tasks

We examined data from eight subjects as they performed or imagined repetitive hand or tongue movements in response to a visual cue. All eight subjects performed the motor movement tasks, and six of the subjects also performed the motor imagery task.

Thirty 3-s visual word cues for hand and tongue movement were presented interleaved in random order with a 3-s rest period between each cue. The cues were delivered in a 10×10-cm presentation window at a distance of 75–100 cm from the subject. In response to each stimulus, the subject performed a repetitive movement of the hand (clenching and unclenching) or tongue (sticking out the tongue) for the period of the stimulus presentation. In a separate session, the subjects imagined these movements in response to the stimulus, without performing physical movements.

Our study used repetitive motion, rather than tonic contraction, in order to accentuate the spectral shift during each interval, as attenuation of alpha and beta ERD [14] and gamma ERS [15] has been reported during tonic contraction.

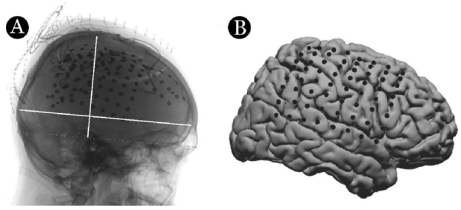


Fig. 1. Example of ECoG electrode grid locations: Part A shows an X-ray image of the patient's brain with a grid and strip electrodes implanted on the surface of the brain. The white lines are used to estimate 3-D electrode locations from the image, in conjunction with other X-ray images. Part B shows the electrode locations as placed on a 3-D computer model of a standardized brain.

Due to clinical circumstances, we cannot collect large amounts of training data, and our subjects' compliance with the experimental protocol is likely to be variable across subjects and during sessions.

III. Feature Representation

We transform a window of data from each channel into two features: 1) the lowband power (11–40 Hz) and 2) the highband power (71–100 Hz) features (see Fig. 2). The features are calculated as the log variance of the specified window of data from a channel after it has been bandpass-filtered in the appropriate band. The bands are chosen in order to exclude the possibility of 60-Hz line noise artifacts.

Fig. 2 shows two fairly typical examples of spectral power changes associated with motor activity or imagery. The two electrodes marked in the figure are from the hand and tongue areas, respectively, and show that, broadly speaking, there is a decrease in spectral power in the lowband feature and an increase in spectral power for the highband feature during movement.

Our choice of feature selection was motivated by two compelling reasons: 1) We have consistently seen quantitative differences in these bands between the average spectra for movement and rest across subjects and motor actions, showing that this is a general physiological phenomenon. Recent work [16] has shown that this spectral change can be reliably used in place of electrical stimulation to localize motor representations in the brain. 2) The paucity of data (only 30 trials per class, for up to 100 channels) forces us to use a single, simple set of features across all subjects in order to prevent overfitting. Finally, Section V shows, in post-hoc analysis, that the feature weighting chosen by the classification methods also shows the characteristic lowband suppression and highband increase in activity that is described here.

Individual differences in bandpower feature modulation certainly do exist in our subject populations, and our reported results may benefit from additional training data that can be used to customize the bandpower features to each individual subject.

IV. Classification Methods

We explored four classification methods in our experiments: the (regularized) linear discriminant analysis (LDA) classifier, the support vector machine (SVM), and sparse variants of these two methods. All four methods used are linear binary classifiers (i.e., they assume linear separability of two classes of data and

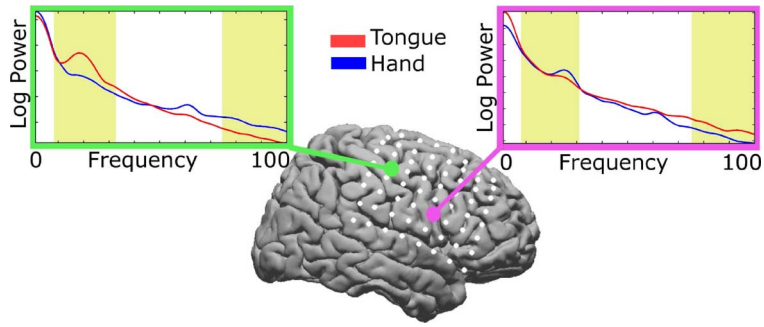


Fig. 2. Spectral features: The figure illustrates our choice of spectral features. Shown are average spectra during tongue and hand movement tasks for two channels taken from the classical hand and tongue cortical representation areas. We see that for electrodes over the relevant body part representation, activity in that body part produces suppression in the lowband power and an increase in the highband feature.

attempt to find a hyperplane separating the data points belonging to the two classes). We first describe the binary classifiers in detail and subsequently describe how multiclass problems can be solved by combining binary classifiers.

A linear classifier is represented by $(\mathbf{w}, b)$ which are the normal to the separating hyperplane, and its distance from the origin, respectively. The classifier is used to compute the label $y \in \{+1, -1\}$ of any given point $\mathbf{x}$ as $\mathbf{y} = \text{sign}(\mathbf{w}^T\mathbf{x} + b)$. Linear classifiers have two powerful advantages: simplicity and interpretability. For example, since the classifier output is a weighted linear combination of all input features, we can examine the components of the weight vector $\mathbf{w}$ to see which features are considered important by the classifier. Also, in problems with limited training data (as in our case), they help alleviate the risk of overfitting.

Sparse linear classifiers seek a sparsely populated projection vector $\mathbf{w}$ (i.e., a weight vector with most components at zero or close to zero). Typically, these sparse methods allow a tradeoff between sparseness of $\mathbf{w}$ and training set error. This allows us to automatically discover and use only the most important features in the input vector $\mathbf{x}$. Thus, sparse classifiers are useful when the input data has a large number of irrelevant features.

A. Regularized Linear Discriminant Analysis (LDA)

LDA, also known as the Fisher's linear discriminant, is a simple statistical approach to separating data from multiple classes, and is commonly used for classification, feature extraction, and dimensionality reduction. LDA (see, for example, [17] for details) chooses a projection vector $\mathbf{w}$ that maximizes the separation between the projected means of the two classes. This direction is computed using the class means μ_1, μ_2, and the within-class scatter, S_w. The within-class scatter is the sum of the *a priori* probability of each class times the covariance of that class

$$S_w = \sum_c p_c \times \text{cov}_c.$$

In the two-class case, we take $p_1 = p_2 = 0.5$. Using the within-class scatter matrix S_w, we define $\mathbf{w}$ and the offset b as

$$\begin{aligned} \mathbf{w} &= \text{inv}(S_w) \times (\mu_2 - \mu_1)^T \\ b &= -\frac{1}{2}(\mu_1 + \mu_2) \times \mathbf{w}. \end{aligned} \tag{1}$$

The traditional LDA classifier contains an implicit constraint in that it requires S_w to be invertible and, hence, nonsingular. In our case, the sample size is significantly smaller than the dimensionality of the data, and this constraint is often not satisfied. RLDA is a simple variant of LDA that regularizes the scatter matrix S_w by adding constant values to the diagonal elements, thereby guaranteeing the nonsingularity of S_w. For a choice of parameter value $0 \leq \lambda \leq 1$, the regularized scatter matrix is given by

$$S_w \leftarrow (1 - \lambda)S_w + \lambda I_m. \tag{2}$$

As $\lambda \to 1$, the information in S_w is lost and as $\lambda \to 0$, the regularization term is discarded. The parameter λ is a free parameter chosen via model selection to minimize generalization error.

B. Support Vector Machines (SVMs)

The SVM classifier [18] chooses the hyperplane of maximum margin, or "thickness" that separates the two classes of data. This choice is expected to be more robust to outliers in the data, and SVMs have been popular and successful in a wide variety of applications. It can be shown that the margin width of the separating hyperplane is inversely proportional to $\|\mathbf{w}\|_2^2$. Here, we use $\|\cdot\|_2$ to represent the Euclidean or l_2-norm, and $\|\cdot\|_1$ for the l_1-norm (i.e., $\|\mathbf{w}\|_1 = \sum|\mathbf{w}|$). The search for the optimal $\mathbf{w}$ can thus be framed as a quadratic optimization problem, subject to the constraints that each training data point is correctly classified. Further, to allow for misclassifications and outliers, slack variables ξ_k are introduced. Thus, if the m-dimensional data samples are $\mathbf{x}_k, k = 1, \ldots, K$, where K is the number of samples, and the class membership with $y_k \in \{-1, +1\}$, then the optimization problem for the SVM is

$$\begin{aligned} \min_{\mathbf{w},\xi,b} \quad & \frac{1}{2}\|\mathbf{w}\|_2^2 + \frac{C}{K}\|\xi\|_1 \quad \text{subject to} \\ & y_k(\mathbf{w}^T\mathbf{x}_k + b) \geq 1 - \xi_k \quad \text{and} \\ & \xi_k \geq 0, \quad \text{for} \quad k = 1, \ldots, K. \end{aligned} \tag{3}$$

The constraints ensure that the training data are correctly classified, and the ξ_k terms are used as a regularization term and allow for errors, with C being a free parameter that controls the regularizer.

C. Sparse Classification Methods

The standard linear Fisher's discriminant can be recast [19] as the solution to the following quadratic optimization problem:

$$\min_{\mathbf{w},\xi,b} \quad \frac{1}{2}\|\mathbf{w}\|_2^2 + \frac{C}{K}\|\xi\|_2^2 \quad \text{subject to}$$
$$y_k(\mathbf{w}^T\mathbf{x}_k + b) = 1 - \xi_k, \quad \text{for} \quad k = 1, \ldots, K. \quad (4)$$

This formulation is very similar to the quadratic program used for the SVM. Both of these quadratic programs can be converted to linear programming problems by replacing the l_2-norm on the regularizer with the l_1-norm. In the case of the Fisher's discriminant, the l_1-norm is also used on the slack variable ξ. An additional advantage to sparseness is that the resulting linear programs are simpler to solve than the original quadratic programming counterparts.

Thus, we have two sparse classification methods—the linear programming machine (LPM), and the linear sparse fisher's discriminant (LSFD). The LPM is the solution to the following optimization problem:

$$\min_{\mathbf{w},\xi,b} \quad \frac{1}{N}\|\mathbf{w}\|_1 + \frac{C}{K}\|\xi\|_1 \quad \text{subject to}$$
$$y_k(\mathbf{w}^T\mathbf{x}_k + b) \geq 1 - \xi_k \quad \text{and}$$
$$\xi_k \geq 0, \quad \text{for} \quad k = 1, \ldots, K. \quad (5)$$

The LSFD is the solution to the following optimization problem:

$$\min_{\mathbf{w},\xi,b} \quad \frac{1}{N}\|\mathbf{w}\|_1 + \frac{C}{K}\|\xi\|_1 \quad \text{subject to}$$
$$y_k(\mathbf{w}^T\mathbf{x}_k + b) = 1 - \xi_k, \quad \text{for} \quad k = 1, \ldots, K. \quad (6)$$

While the two sparse methods are solutions to very similar optimization problems, they are conceptually addressing different goals: maximizing the margin between the two classes (LPM), and maximizing the distance between the two class means while minimizing variance along the projection dimension (LSFD).

Also, in these two methods, the free parameter C now controls the tradeoff between sparsity of the weight vector and the errors made by the classifier—a high value of C would impose a more severe penalty on misclassifications, and a lower C value would favor sparseness of the weight vector $\mathbf{w}$. This parameter is selected empirically using model selection.

D. Model Selection and Evaluation

Each classifier has a free parameter that is chosen empirically to minimize generalization error. This is accomplished by evaluating cross-validation accuracy of the classifier on a given dataset data for a range of parameter values, and choosing the parameter value that minimizes cross-validation error. In order to test the generalization of the classifier with the chosen parameter value, we use double-crossvalidation as our performance measure. Specifically, we randomly divide the trials into five blocks, using four for training and one for testing. In each train step, we select classifier parameters by minimizing five-fold cross-validation error on the training data. We then use all of the training data with the selected parameter value to train the classifier and evaluate its performance on the test data. Thus, the classifier is tested on data points that are unseen during the training and parameter selection phase. The entire nested cross-validation routine is repeated ten times and the average error over all runs is presented as a measure of classifier performance.

We implemented our classifiers with the use of Matlab's linprog linear optimization package. For SVM, we used the LIBSVM package [20].

V. Results

A. Classification Error

Fig. 3 presents the classification error of each method on each subject for both real and imagined movements. For this experiment, windows of data from 1–3 s in each trial were converted to bandpower features, thus yielding 30 trials. Even with so little data, the motor action classification results (average 6% error for LPM, including one outlier) are better than previously reported EEG results (e.g., the Berlin BCI [3]) where the best reported results are in the 10% range. The motor imagery results are comparatively worse (average 23% error for the LPM classifier), but are still comparable to previous ECoG results (e.g., Lal *et al.* [7] who used 100–200 samples in three subjects to obtain errors of 17–23%). The interesting feature about the motor imagery results is the high amount of intersubject variance, where we believe subject compliance is an issue.

In a recent initial study [21], we have also shown that for motor movement, individual fingers of one hand can be classified with 77% accuracy in a five-class classification problem, indicating that ECoG may indeed have substantially more information about movements compared to EEG.

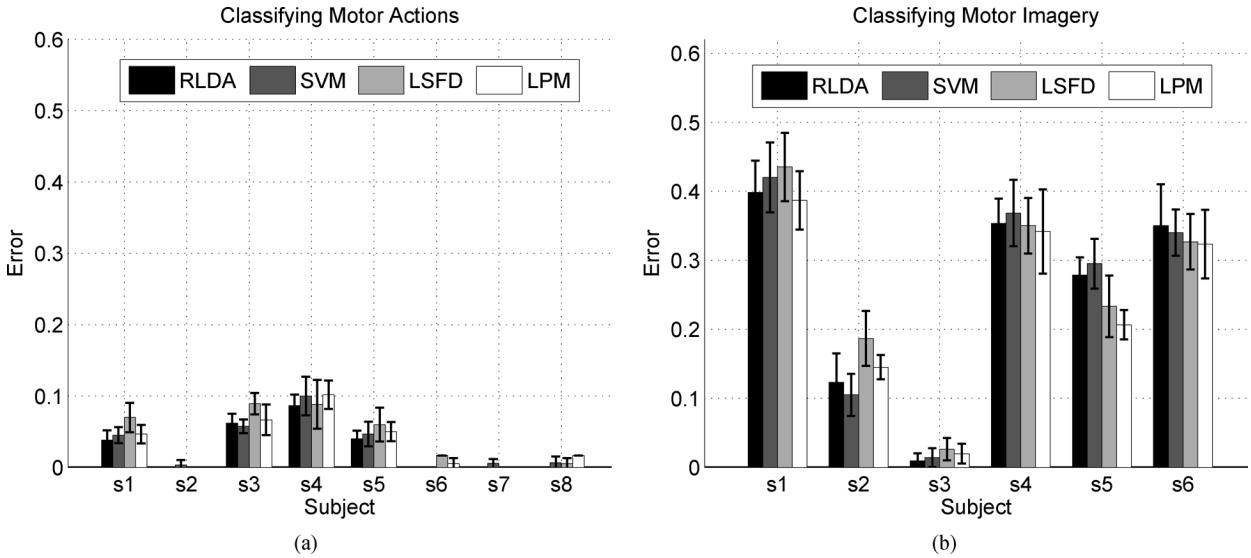
B. Spatial Features

We examined the weights chosen by each classifier to see which spatial features were selected by the classifier. In order to understand the spatial selectivity in aggregate, we normalized each subject's classifier weights to unit length and projected the weights onto a standard brain using estimated electrode positions. The electrode positions in Talairach standardized coordinates [22] were calculated using anterior-posterior and lateral skull X-rays [23].

Fig. 4 shows the cumulative projection of all subjects' weight vectors onto the brain. These figures were created by scaling and linearly superimposing spherical Gaussian kernels (width 5-mm variance 25 mm) centered at the location of each electrode on a template brain.

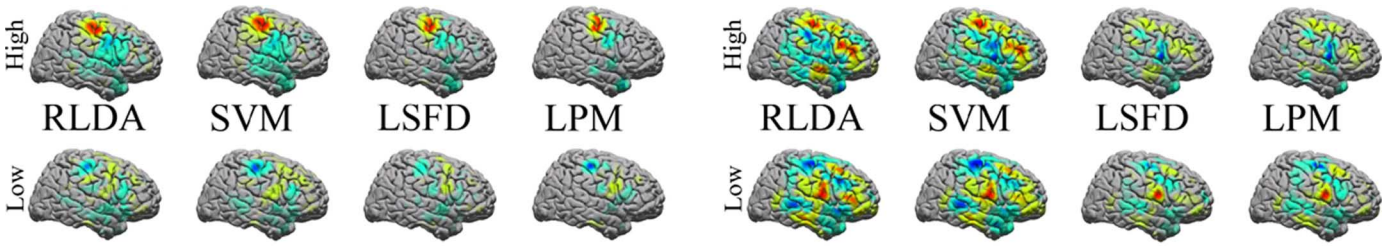
The figures clearly indicate spatial clustering of the selected features across subjects, showing that the classifiers have automatically extracted generalizable spatial features which correspond to known somatotopic locations. This result also indirectly supports our argument for choosing a single, general set of spectral features, since the low and high spatial features chosen by each classifier are spatially similar and opposite in sign (see Section III).

Not suprisingly, the sparse methods select more spatially focused features compared to the other two methods, especially for the motor imagery data. This useful property may help reduce the problem of overfitting, as well as reduce the number of channels required online for control.



(a) (b)

Fig. 3. Classification Error: This figure shows the double-cross-validation error of the classifiers on all datasets. Figure (a) shows the results for all eight subjects during overt motor actions, and (b) shows results for the six subjects who also participated in the motor imagery task.



(a) (b)

Fig. 4. Spatial features for motor action and imagery: The weight vectors for each subject are plotted onto the brain in separate low-feature and high-feature plots per classifier. Each column in figure (a) shows the low- and high-frequency features selected by the methods for motor action data. The sparse methods can be seen to select spatially more focused channels. Figure (b) shows similar results for motor imagery.

C. Using Fewer Features

We saw in the previous sections that while the average error for the different methods does not differ significantly, the sparse methods select a highly focussed feature set for classification. In this experiment, we attempt to quantify this property of the sparse methods. Specifically, we train the classifiers on the training data exactly as before (in a nested cross-validation routine), but subsequently select a fraction, say 20%, of the weights in the classifier's projection vector $\mathbf{w}$. The weights with the largest absolute magnitude are selected, and the other components of $\mathbf{w}$ are zeroed out. This trimmed weight vector is then used to classify the unseen test data.

Note that this is not a completely fair evaluation, since the classifier optimization criterion does not include this additional constraint. Nevertheless, the outcome of this experiment is informative. Figs. 5 and 6 show the results for motor action and motor imagery, respectively, averaged over all subjects. In each pair, the first plot shows the distribution of weights in the weight vector, normalized by the largest value. For example, Fig. 5(a) indicates that for the LPM classifier, on average more than half the weights are zero while, for the nonsparse methods, all of the weights are up to 40% of the largest weight. The second plot shows the classifier performance when using only the weights with the largest magnitude. We see that for the LPM classifier, the error drops steeply, and is constant after the top 20% components are used. This is strong evidence that for the sparse methods, the magnitude of the weights for each individual feature can be used for ranking them in order of importance and used directly for classification.

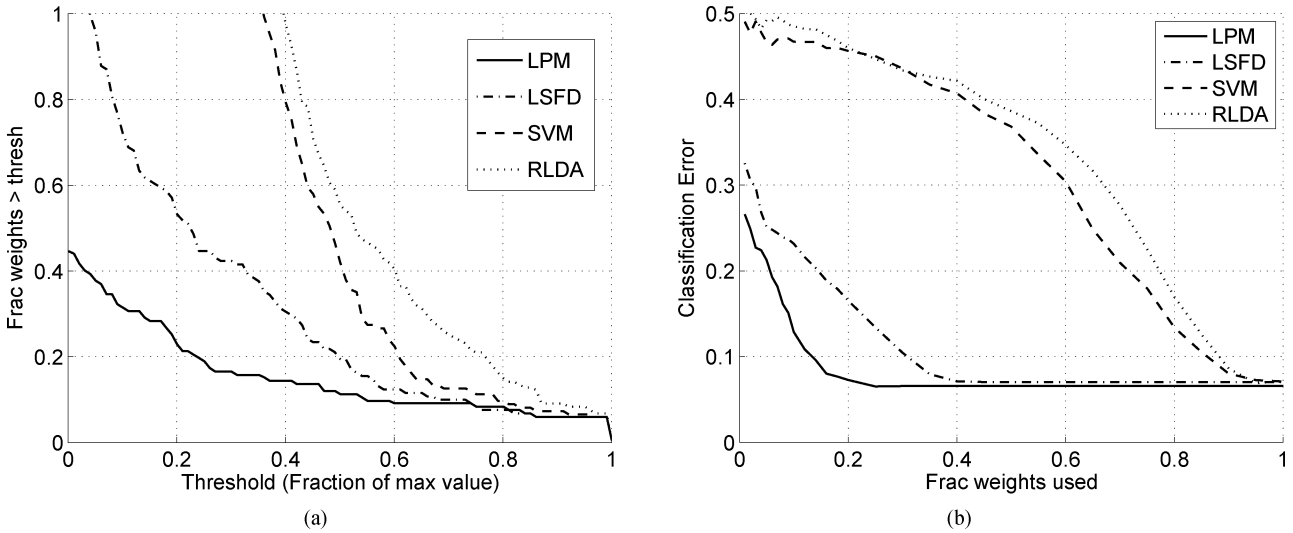
It is conceivable that the other methods also contain ranking information in the weight magnitudes. However, an additional thresholding and retraining step would be required to exploit this information (e.g., as in recursive channel elimination [7]). Given the paucity of data, we refrain from attempting a second training step and evaluation.

Fig. 6 shows similar results for motor imagery. Interestingly, even though, on average, a greater number of components are nonzero for the sparse methods, the average error still drops to its lowest with the use of around 20% of the features.

D. Leveraging Data From Overt Movements

The very high classifiability of overt movement data, combined with the focal nature of the selected features, naturally leads one to consider the following question: Can we use the features selected for classifying overt movements to improve classification performance on the imagery data?

To test this idea, we used the magnitude of the weights from the learned classifiers as a score for channels. We then selected the top 20% channels, ranked by magnitude, for each classifier



(a) (b)

Fig. 5. Fewer features for motor action. (a) The distribution of weight magnitudes for each classifier is shown. The sparse classifiers have on average a large number of zeros, whereas the nonsparse methods have very large weight components. (b) The error when only the components with the highest magnitudes are retained in the classifier is shown. The error for the sparse methods drops steeply, and reaches the minimum with approximately 20% of the features being used.

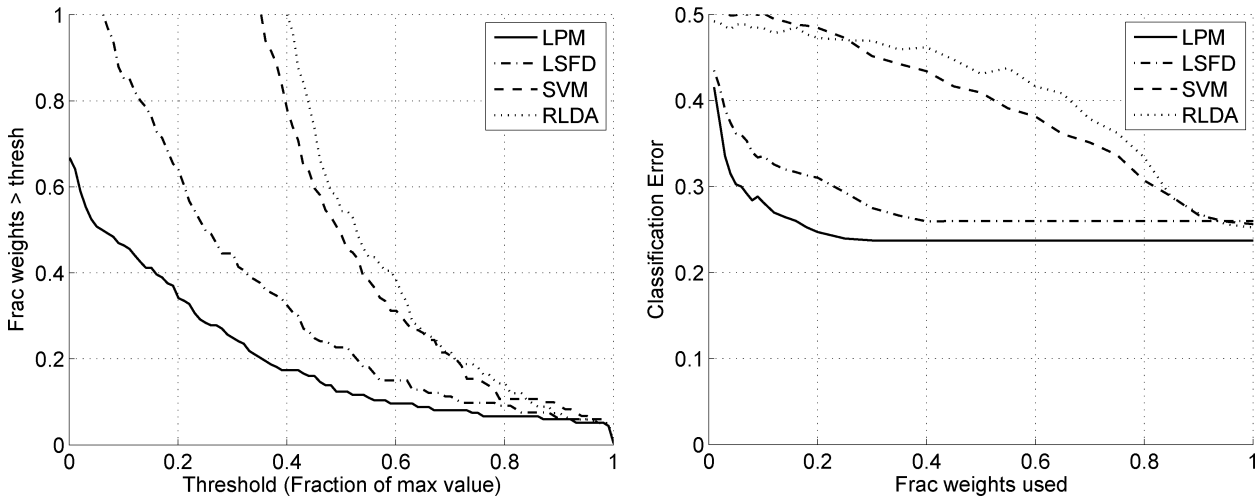


Fig. 6. Fewer features for motor imagery. Analysis of weight magnitudes for the different classifiers (see Fig. 5).

and subject. The imagery data were then restricted to only those selected channels, and the methods were again evaluated using double-cross-validation on this restricted set of channels.

We see in Fig. 7 that this method indeed does improve the performance (cf. Fig. 3). This is an important result since only a fifth of the channels are used for classification in this experiment. Surprisingly, the nonsparse methods—SVM and RLDA—improve the most, on average, with a 5% and 4% reduction in error, respectively. The sparse methods do not show an overall trend in error, with improvements in some datasets and decrease in performance for other subjects. One possibility is that after restricting the channels to only the relevant motor channels, the sparsity constraint is a severe penalty. This result also supports the claim that information is indeed encoded in the feature weights chosen by the SVM (cf. the RCE method used by Lal *et al.* [7]).

In the ideal case of infinite data, this step should not be necessary (i.e., the classifier should automatically select relevant features for motor imagery). However, given the small amount of high dimensional training data, and the fact that classifying motor imagery data is significantly harder, a preprocessing step that identifies neurophysiologically relevant channels improves performance in practice. Since the selection is made on a different dataset, overfitting is unlikely.

This method can only be directly used in subjects who still retain some degree of motor control, and may not be applicable to paralyzed or already incapacitated individuals. However, the similarity of real and imagined movements that is implied by these results allow us to use algorithms that are designed and tested on real movements by healthy subjects as a starting point for incapacitated users.

E. Discussion

We see from the results that ECoG signals recorded during motor action are highly separable, with very low average errors. In contrast, the classification performance on motor imagery data is harder and more variable, with errors ranging from 2% to 30%, and an average error of 20%. While these errors

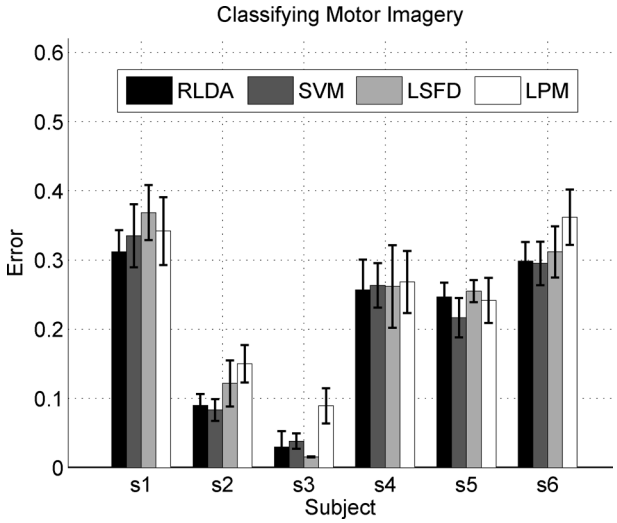


Fig. 7. Using action features for imagery data. For each method, the top 20% channels from the motor action task (chosen by weight magnitude) are used to train and test the classifiers on the motor imagery data. The SVM classifier error improves by 5% on average.

are comparable to previous ECoG classification results (e.g., Lal *et al.* [7], who obtain 17.5%–23.3% on three subjects, with 100–200 data points for training), they are significantly worse than classifying motor action. In this regard, we make the following observations: First, only 30 samples of data per class were used, an amount considerably smaller than that used in typical BCI studies. Second, the neurophysiologically relevant hand and tongue areas are also weighed heavily in the motor imagery task. Third, restricting the channels to those ranked highest for the motor action task significantly improved prediction performance on imagery data. Fourth, patient compliance is difficult to guarantee in a motor imagery task. These observations indicate that more data are likely to improve performance. Further studies would be needed to confirm this hypothesis and discover the limits of the ECoG signal for BCI.

We explored four classification methods, and a new generalized spectral feature representation for classifying ECoG signals. Our choice of spectral features was strongly supported by the posthoc analysis of the feature weightings chosen by the different classifiers (see Fig. 4), where the neurophysiologically relevant channels were ranked the highest, and the LPM classifier selected spatial features that were highly focused on these areas. Our results indicate that while all methods performed comparably, sparse classification methods used a much smaller set of features for very good classification performance. In addition, the magnitude of the weights can be used as an explicit ranking of the feature quality, allowing us to restrict the number of features used. We also showed that using information learned from motor actions for interpreting data from motor imagery is both feasible and beneficial. This fact suggests that similar spatial areas are involved in ECoG changes during motor action and motor imagery.

Our study forms a first step toward exploring the usability of ECoG signals as a BCI input mechanism. This paper shows that, in contrast to EEG, motor movements can be classified with high accuracy, using very little training data. Subsequent work [21] has shown that very fine distinctions can be made using ECoG, including the classification of individual fingers of one hand in a five-class classification setting. This level of information is not available on a single-trial basis using EEG. However, the invasive nature of the ECoG recording procedure means that a substantial further study, especially long-term implantation, may be required before ECoG becomes a viable method for brain-actuated control.

VI. Conclusion

We examined classifiability of ECoG signals for use in a minimally invasive human brain–computer Interface. We showed that across eight subjects, the same spectral features and spatial features are involved in motor actions, and these features closely correspond to the underlying neurophysiology of the motor activity. Data from motor actions are highly classifiable, with an average error of about 5%, and motor imagery is classifiable with an average of 20%.

Our comparison of sparse and nonsparse classification methods indicates that in the scenario of minimal training data, the sparse methods may provide significant benefits in terms of interpretability and noise rejection. They are also very useful as automatic feature selection methods.

Future work includes testing classification-based methods in an online feedback scenario, examining the very quick learning and adaptation that takes place in the brain in ECoG-based BCIs [10] and exploring multiclass BCIs.

References

[1] J. Wolpaw, N. Birbaumer, D. McFarland, G. Pfurtscheller, and T. Vaughan, "Brain-computer interfaces for communication and control," *Clin. Neurophys.*, vol. 113, pp. 767–791, 2002.

[2] J. Wolpaw and D. McFarland, "Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans," *Proc. Nat. Acad. Sci.*, vol. 101, no. 51, pp. 17849–54, 2004.

[3] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Mueller, V. Kunzmann, F. Losch, and G. Curio, "The Berlin brain-computer interface: EEG-based communication without subject training," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 14, no. 2, pp. 147–152, Jun. 2006.

[4] D. Taylor, S. Tillery, and A. Schwartz, "Direct cortical control of 3d neuroprosthetic devices," *Sci.*, vol. 296, pp. 1829–1832, 2002.

[5] W. Wu *et al.*, "Neural decoding of cursor motion using a Kalman filter," *Advances NIPS 15*, 2003.

[6] L. Hochberg, M. Serruya, G. Friehs, J. Mukand, M. Saleh, A. Caplan, A. Branner, D. Chen, R. Penn, and J. Donoghue, "Neuronal ensemble control of prosthetic devices by a human with tetraplegia," *Nature*, vol. 442, 2006.

[7] T. N. Lal, T. Hinterberger, G. Widman, M. Schröder, J. Hill, W. Rosenstiel, C. E. Elger, B. Schölkopf, and N. Birbaumer, "Methods towards invasive human brain computer interfaces," *Advances NIPS*, vol. 17, 2005.

[8] B. Graimann, J. Huggins, S. Levine, and G. Pfurtscheller, "Towards a direct brain interface based on human subdural recordings and wavelet packet analysis," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 954–962, Jun. 2004.

[9] E. Leuthardt, G. Schalk, J. Wolpaw, J. Ojemann, and D. Moran, "A brain-computer interface using electrocorticographic signals in humans," *J. Neural Eng.*, vol. 1, no. 2, pp. 63–71, 2004.

[10] E. Leuthardt, K. Miller, G. Schalk, R. Rao, and J. Ojemann, "Electrocorticography-based brain computer interface-the seattle experience," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 14, no. 2, pp. 194–198, Jun. 2006.

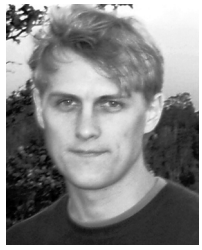
[11] G. Pfurtscheller and F. L. da Silva, *Handbook of Electroencephalography and Clinical Neurophysiology – Event-Related Desynchronization*. Amsterdam, The Netherlands: Elsevier, 1999.

[12] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, and K.-R. Mueller, "Optimizing spatio-temporal filters for improving brain-computer interfacing," *Advances NIPS*, vol. 18, 2006.

[13] G. Schalk, D. McFarland, T. Hinterberger, N. Birbaumer, and J. Wolpaw, "BCI2000: A general-purpose brain-computer interface (BCI) system," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1034–1043, Jun. 2004.
[14] N. Crone *et al.*, "Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. I. Alpha and beta event-related desynchronization," *Brain*, vol. 221, pp. 2271–99, 1998.
[15] N. Crone *et al.*, "Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. II. Event-related synchronization in the gamma band," *Brain*, vol. 221, pp. 2301–15, 1998.
[16] K. Miller *et al.*, "Electrocorticographic spectral changes with motor movement," *J Neurosci.*, 2007, to be published.
[17] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2000.
[18] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer Verlag, 1995.
[19] S. Mika, G. Rätsch, and K. Müller, "A mathematical programming approach to the kernel Fisher algorithm," *Advances in NIPS*, vol. 13, p. 591, 2001.
[20] C.-C. Chang and C.-J. Lin, LIBSVM: A Library for Support Vector Machines 2001. [Online]. Available: http://www.csie.ntu.edu.tw/ cjlin/ libsvm.
[21] P. Shenoy, K. Miller, J. Ojemann, and R. Rao, "Finger movement classification for an electrocorticographic BCI," presented at the IEEE Neur Eng. Conf. 2007.
[22] J. Talairach and P. Tournoux, "Talairach: Co-planar stereotaxic atlas of the human brain," *Stuttgart: Thieme*, 1988.
[23] K. Miller *et al.*, "Cortical surface localization from x-ray and simple mapping for electrocorticographic research: The location-on-cortex package," *J. Neurosci. Methods*, 2007.

Pradeep Shenoy received the Master's degree in computer science from the University of Washington, Seattle, in 2004.

His research interests include the application of machine learning techniques to the understanding of EEG and ECoG signals, with an emphasis on applications in brain–computer interfaces.



Kai J. Miller is currently pursuing the M.D./Ph.D. degree at the University of Washington, Seattle.

He is with the Departments of Physics, Computer Science, and Medicine at the University of Washington. His primary interest is the investigation of computational strategies employed by the nervous system in order to affect behavior. He has worked on the encoding of movement-related information in ECoG signals, and their application to localizing cortical representations of movement.

Jeffrey G. Ojemann is an Associate Professor of neurological surgery at the University of Washington, Seattle, and holds the Richard G. Ellenbogen Chair in Pediatric Neurological Surgery. His work looks at electrical signals recorded directly from the surface of the brain (electrocorticography) in patients undergoing treatment of epilepsy. He investigates behavioral correlates of ECoG signal changes, and the fundamental properties of these signals, along with collaborators in computer science and physics. His recent work includes the effective application of these insights in brain–computer interfaces.

Rajesh P. N. Rao is an Associate Professor in the Computer Science and Engineering Department at the University of Washington, Seattle, where he heads the Laboratory for Neural Systems. Rao is the co-editor of two books: Probabilistic Models of the Brain (2002) and Bayesian Brain (2007).

Prof. Rao is the recipient of a David and Lucile Packard Fellowship, an Alfred P. Sloan Fellowship, an Office of Naval Research Young Investigator Award, and a National Science Foundation Career award.