Journal of Neural Engineering

Autoencoders for learning template spectrograms in electrocorticographic signals

Autoencoders for learning template spectrograms in electrocorticographic signals

Tejaswy Pailla[1], Kai J. Miller[2], Vikash Gilja[1]

[1] Department of Electrical and Computer Engineering, University of California-San Diego, San Diego, California, United States of America
[2] Department of Neurosurgery, Stanford University School of Medicine, Stanford, California, United States of America

E-mail: tpailla@eng.ucsd.edu

March 2018

Abstract.
Objective: Electrocorticography (ECoG) based studies generally analyze features from specific frequency bands selected by manual evaluation of spectral power. However, the definition of these features can vary across subjects, cortical areas, tasks and across time for a given subject. We propose an autoencoder based approach for summarizing ECoG data with "template spectrograms", i.e. informative time-frequency (t-f) patterns, and demonstrate their efficacy in two contexts: brain-computer interfaces (BCIs) and functional brain mapping. *Approach*: We use a publicly available dataset wherein subjects perform a finger flexion task in response to a visual cue. We train autoencoders to learn t-f patterns and use them in a deep neural network to decode finger flexions. Additionally, we propose and evaluate an unsupervised method for clustering electrode channels based on their aggregated activity. *Main result*: We show that the learnt t-f patterns can be used to classify individual finger movements with consisentently higher accuracy than with traditional spectral features. Furthermore, electrodes within automatically generated clusters tend to demonstrate functionally similar activity. *Significance*: With increasing interest in and active development towards higher spatial resolution ECoG, along with the availability of large scale datasets from epilepsy monitoring units, there is an opportunity to develop automated and scalable unsupervised methods to learn effective summaries of spatial, temporal and frequency patterns in these data. The proposed methods reduce the effort required by neural engineers to develop effective features for BCI decoders. The clustering approach has applications in functional mapping studies for identifying brain regions associated with behavioral changes.

Keywords: brain-machine interfaces, functional mapping, time-frequency analysis, electrocorticography, neural signal processing
Submitted to: *J. Neural Eng.*

1. Introduction

Electrocorticography (ECoG) is a neural recording technique wherein electrical potentials are measured from the surface of cortex. One of the primary clinical applications of ECoG is to evaluate epileptic activity: individuals with chronic intractable epilepsy are implanted with ECoG electrodes and their cortical activity is monitored in an epilepsy monitoring unit (EMU) to localize epileptic seizure foci. In this setting, the implanted electrodes are also used to provide electrical stimulation to map functional brain areas. These maps guide surgical resection of affected brain areas while attempting to minimize the potential for functional deficits. Patients in the EMU are often recruited to take part in neurophysiology studies and ECoG data are collected while they perform cognitive and behavioral tasks. Given this unique opportunity to study the human brain in action, such studies have significantly contributed to the advancement of human neuroscience research, particularly with respect to two domains: functional brain mapping [1, 2] and brain-computer interface (BCI) development [3].

Most functional brain mapping studies analyze spatio-temporal patterns of ECoG signal modulation in the 8-40 Hz band or the 70-200 Hz band or sub-bands in those ranges to identify cortical processes underlying behaviors [2, 4, 5, 6, 7]. Precise labeling of behavioral onset and offset timings is important when performing such analyses since event related potentials are sensitive to the timing of activity [8]. However, many of the labeling methods employed in experimental studies rely upon the subjects' reaction times to a cue and this can potentially introduce errors in analysis; thus, while ECoG may afford high temporal resolution, study design and conventional analysis methods can limit the temporal resolution of interpretations.

The high spatio-temporal resolution and signal-to-noise ratio (SNR) of ECoG also makes it a promising source for BCI control signals. Several studies have explored the use of ECoG for prosthetic control [3, 9, 10, 11, 12] and ECoG based BCI systems are being explored for long term use as communication prosthesis for individuals with locked-in syndrome [13]. Evidence of ECoG signal modulation in specific spectral bands, such as 8-32 Hz and 76-100 Hz [14] and in broad spectral changes in the 5-200 Hz frequency range [15] during motor activity, has motivated the majority of ECoG based BCI studies to use frequency specific spectral power as features for decoding. These spectral powers are typically estimated from a specific time window around activity onset. However, the typically applied conventional and rigid feature selection of frequency bands and time windows may limit BCI system performance, since the precise range of informative frequency bands and temporal range of relevant modulation may vary across subjects [16]. Also, the activation patterns in different cortical areas for the same task might vary for the same subject throughout the duration of the experiment as the subject's familiarity with the task changes [17]. Furthermore, computing spectral powers in aggregated time bins does not fully exploit the temporal resolution that ECoG provides. With conventional approaches, decoding ECoG signals with high accuracy requires a neural engineer to carefully analyze spatial, temporal and frequency patterns in the data and hand design custom subject-specific, task-specific features and decoding models, resulting in a herculean task.

Thus, the complexity of feature engineering motivates the need for automated pattern recognition tools that can efficiently extract spatial, temporal and frequency patterns in ECoG data. In recent years, several neuroimaging and EEG based studies have used Deep Learning (DL) methods for pattern recognition and classification tasks [18, 19, 20, 21, 22, 23]. Deep belief networks (DBN) have been applied to learn neural correlates in real-time for an ECoG based BCI[24]. In another study [25], a variant of convolutional stacked autoencoder [26] has been used to extract hierarchical features from ECoG signals. Other studies [27, 28, 29, 30] have used echo state network implementations of recurrent neural networks (RNNs) to model neural activity for use in BCIs. These studies demonstrate that deep neural network architectures can achieve performance comparable to the state-of-the-art achieved by conventional machine learning methods and, in addition, have the potential to extract physiologically relevant features [18]. However, most studies that use deep learning methods to decode neural data use network architectures that are successful in other domains, mainly computer vision (For example, [31] used a network mimicking the VGG architecture that was developed for object recognition). Although such architectures achieve good performance results, the number of parameters to be learnt is on the order of millions, thus risking overfitting. Given the small dataset sizes in neurophysiology studies, network architectures designed with domain knowledge could address the problem of overfitting and improve generalizability of architectures across subjects and tasks.

In this work, we learn generalized time-frequency (t-f) patterns, which we call "template spectrograms" (TS), that can effectively encode patterns from ECoG electrode channels at different spatial locations using autoencoders (AE) [32]. These patterns find representations that are robust to temporal and frequency shifts in the data. We show the efficacy of TS in two contexts: BCI decoding and identification of functionally similar cortical regions (Figure 1). We use a convolutional neural network framework that uses TS to decode individual finger flexions, i.e. identify which finger is moving, from ECoG data and demonstrate classification performance that surpasses models that achieved when using traditional spectral features. We also evaluate the generalizability of TS by using a subject-to-subject transfer learning approach. The results suggest that for this finger flexion task, informative neural patterns across brain regions and subjects can be summarized using a limited number of TS. To investigate potential functional mapping applications, we use learnt t-f representations to cluster functionally similar brain regions and validate clustering results by comparing them with prior knowledge of electrode location. We observed that brain areas that have task relevant responses are clustered together. This suggests that unsupervised learning based methods could aid in the mapping of brain areas to behaviors.

2. Methods

2.1. Recordings

Data were collected from 9 subjects with intractable epilepsy who underwent temporary placement of subdural electrode arrays to localize seizure foci prior to surgical resection. All the subjects participated in a purely voluntary manner, after providing informed written consent, under experimental protocols approved by the Institutional Review Board of the University of Washington (#12193). All patient data were anonymized according to IRB protocol, in accordance with HIPAA mandate. These data originally appeared in [33] and were publicly released in [34]. Three subjects' data were excluded from the main analyses due to inconsistencies in recorded behavior and low trial count (see section 2 of Supplementary material for more details about exclusion criteria). Subjects were instructed to move specific individual fingers of the hand contralateral to the implant in response to visual cues. During the finger movement task, subjects were cued with a word displayed on a bedside monitor indicating which finger to move

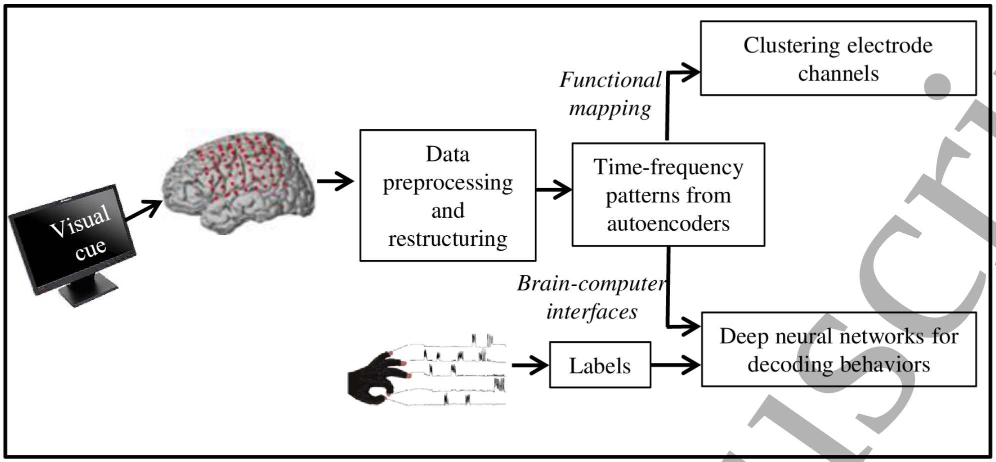
Visual cue
Data preprocessing and restructuring
Time-frequency patterns from autoencoders
Functional mapping
Clustering electrode channels
Brain-computer interfaces
Labels
Deep neural networks for decoding behaviors

Figure 1: **Schematic of proposed electrocorticographic (ECoG) data analysis approach.** ECoG signals and behaviors are synchronously recorded while the subject performs a queued tasks. In the primary datasets considered for this study, a finger flexion task was utilized.

during 2 second movement trials. They performed self-paced movements in response to each of these cues, and typically moved each finger 2-5 times during each trial, but some trials included many more movements. A 2 second rest trial (blank screen) followed each movement trial. There were 30 movement cues for each finger, and trials were interleaved randomly. Finger positions were recorded using a data glove (5DT Data Glove 5 Ultra, Fifth Dimension Technologies; see Supplementary Figures 3 & 4 for example data glove recordings). Electrode locations were determined using localization methods described in [35] and were also manually verified.

2.2. Preprocessing

The ECoG signals were amplified, bandpass filtered between 0.15 and 200 Hz, digitized at 1000 Hz and stored along with the time synchronized digitized flexion samples for all fingers. Noise common to all channels was removed by common average referencing. Signals from each channel were band pass filtered with windowed sinc type I linear phase filters of 4 Hz bandwidth between 4-140 Hz (i.e. 4-8, 8-12 .... 136-140 Hz). After excluding bands that include line noise and its harmonic frequencies (i.e. 56-60, 60-64, 64-68, 116-120, 120-124, 124-128), 28 frequency bands were used. Filtered signals were z-scored. The flexion periods were further broken into 600 ms non overlapping windows which we define as trials for our analysis. Rest data from inter trial intervals were not used. The 600 ms blocks were split into 6 bins of 100 ms each and signal power in each frequency band and each 100 ms time bin was estimated. For all our analysis, to estimate the power of a signal for a frequency band and time bin, we calculated the average of the squared amplitude of the bandpass filtered signal within the time bin. Accordingly, from each electrode channel, trial blocks were restructured into a 28×6 matrix (see Figure 2a) in which each entry is the estimated power for a specific frequency band and time bin. Neural signals were delayed by 150 ms to account for delay between cortical activity and motor movement. Each matrix was labeled with the finger that was flexed during the corresponding block of time, as determined by analyzing the glove data. The method we used for labeling trials is described in section 1 of Supplementary material. The number of 600 ms trials for each subject are 278, 464, 364, 291, 541 and 463, for subjects A-F, respectively.

2.3. Autoencoders for learning time-frequency patterns

An autoencoder, in its simplest form, is a feed-forward neural network that takes an input and maps it to a hidden layer (encoder). The latent representation, i.e. the output from hidden layer is then mapped back (decoder) into a reconstruction of the same shape as input. The weights of encoder and decoder layers are optimized to minimize error between input and reconstructed output [32, 36]. Autoencoders thus learn efficient representation of data and are commonly used for unsupervised feature extraction and dimensionality reduction in neural networks [37, 38].

We use AEs to find t-f patterns (or kernels) in ECoG data. Restructured data from each electrode channel is given as a separate input to an AE and h t-f kernels are learnt using a saturated linear (*satlin*) activation function in Eq. 1, where h is the number of hidden nodes in AE. We expect the *satlin* activation to make the model robust to artifacts by saturating extreme activations that might result from noisy data.

$$f(z) = \begin{cases} 0 & \text{if z} \leq 0 \\ z & \text{if } 0 < \text{z} < 1 \\ 1 & \text{if z} \geq 1 \end{cases} \tag{1}$$

The rationale behind this AE approach is that each channel might have its own task relevant t-f pattern and we want to find a set of templates which have representations generalizable for all channels. To determine the sparsity proportion, i.e. the average output of each hidden unit over the training set, we examined the mean squared error between reconstructed and input signals for varying levels of sparsity proportion and h, and observed that there was no significant trend. So, we used a sparsity proportion of 0.05 for all our analyses. AEs are trained to minimize the sparse mean squared error with l_2 weight regularization for a maximum of 200 epochs. The encoder of a trained AE can be used to generate new t-f based features for neural decoding and in a neural network based decoding architecture (see Section 2.4 and Figure 2b) the encoder can be directly integrated into the architecture. Figure 3a shows example t-f patterns learnt for a subject. We evaluate the efficacy of these representations in two applications: BCI decoding and functional mapping.

2.4. Deep neural networks for decoding

Deep neural networks (DNNs) learn hierarchical representations of data [39] and provide a way to embed feature abstraction as a part of goal-directed learning. In this work, we use a 3-layer AE initialized DNNs (AE-DNNs) to decode individual finger flexions. The architecture is shown in Figure 2b and is implemented with a convolutional framework. N denotes the number of electrode channels for each subject and h_i denotes the number of hidden nodes in each layer. The first layer's weights and biases are fixed with TS learnt from an AE that transforms data from each channel to h_1 nodes. The second layer transforms each of the h_1 nodes in the first layer to h_2 nodes and the last layer is a fully connected softmax layer which maps $N \times h_2$ nodes to 5 nodes that give class probabilities. So, we are effectively learning the discriminative

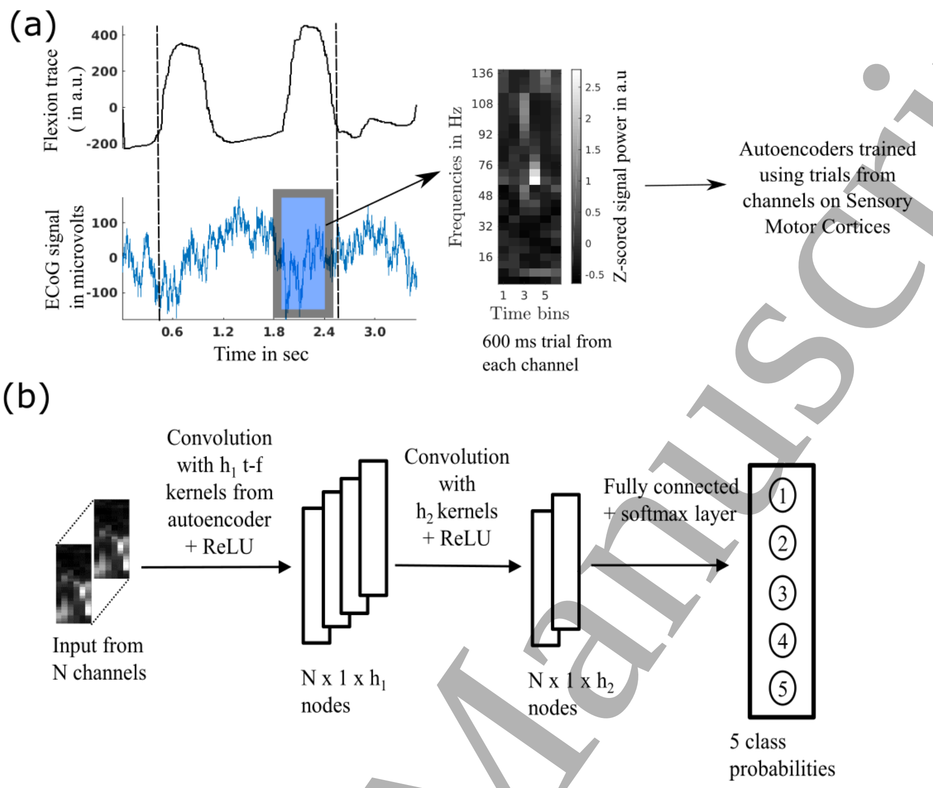
(a)
Flexion trace (in a.u.)
400
200
0
-200
ECoG signal in microvolts
100
0
-100
0.6
1.2
1.8
2.4
3.0
Time in sec
Frequencies in Hz
136
108
92
76
48
32
16
1 3 5
Time bins
2.5
2
1.5
1
0.5
0
-0.5
Z-scored signal power in a.u
600 ms trial from each channel
Autoencoders trained using trials from channels on Sensory Motor Cortices
(b)
Input from N channels
Convolution with h1 t-f kernels from autoencoder + ReLU
N x 1 x h1 nodes
Convolution with h2 kernels + ReLU
N x 1 x h2 nodes
Fully connected + softmax layer
1
2
3
4
5
5 class probabilities

Figure 2: **Schematic of network architecture.** (a): Trials from each channel are restructured in time-frequency (t-f) format and autoencoders are trained to extract t-f patterns. ECoG signal (in blue, shown for a single channel) is synchronized with finger flexion trace (in black). The dashed lines represent the start and end of a flexion period.(b): t-f kernels learnt by autoencoders are used as weights in DNNs to decode finger flexions. DNN takes input from N channels. Weights for each electrode channel, i.e. spatial modeling, are learnt in the last layer.

spatial pattern in the last layer. We use a rectified linear unit (ReLU) non-linearity between the layers as we observed that the training was more stable as compared to training with *sigmoid* and *tanh* non-linearities. Unless otherwise specified, we only train the last two layers of DNN while keeping the first layer weights from AE fixed. We perform 5-fold cross validation (CV) for evaluating decoding performance. The trials are first randomized and DNNs are trained on 4/5th of the data and tested on the remaining 1/5th. While randomizing trials, we ensured that possibly correlated 600 ms trials split from the same flexion epoch are either in training set or in test set. We also ensured that the TS are obtained from AE using only the trials from training set. Section 3 of Supplementary material provides more details about parameter tuning and neural network training. The network with the minimum validation loss is chosen to decode trials in the test set. Sample scripts used for decoding analyses are available at `https://github.com/tejaswy/ECoG-TF`. Neural network training is done with

MatConvNet [40]. For decoding results, we used electrodes from sensory motor cortices (SMC) which includes dorsal M1, dorsal S1 and ventral SMC regions. For each subject, the {number of SMC electrodes}/{total number of electrodes} is 11/46, 11/64, 15/63, 12/47, 11/61 and 9/64 for subjects A-F respectively.

Transfer learning

Limited ECoG dataset size for structured tasks is a major impediment to applying deep learning methods for BCI studies. We show that using a transfer learning (TL) approach, we can leverage the representations learnt across subjects to decode flexions for a new subject. We use data from 4 subjects to learn TS and use them in the first layer of DNN to decode flexions of fifth subject. The rationale behind this is that there may be generalizable t-f patterns in different brain areas across subjects and extracting these patterns from the available data corpus and using them for a new subject by learning the spatial structure by training the last layers might be advantageous in data limited conditions. Thus, the goal for the AE is to effectively learn a "library" of relevant TS.

2.5. Baseline models

Spectral features (SF), i.e. estimated spectral powers in low (8-32 Hz) and high (76-100 Hz) frequency bands are prominently used as features in ECoG based BCIs [14, 41]. Linear Discriminant Analysis (LDA) classifier is commonly used in ECoG based BCI studies [11]. Thus, to evaluate the efficacy of different features, i.e. SF, time-frequency powers described in Section 2.2 (TF), and AE, as well as to evaluate different decoding models, i.e. LDA vs. DNN, we compare decoding results for the following pairings of features and decoding models.

- **SF-LDA**: We use SF calculated in 6 bins of 100 ms each with a regularized LDA classifier. To calculate SF, we bandpass filter signals in 8-32 Hz and 76-100 Hz band and calculate the spectral power estimate using the procedure described in Section 2.2. So, for a given channel, we have 12 SF per trial. This provides a baseline model with traditional SF and a linear decoding model.
- **TF-LDA**: We used estimated powers in time-frequency arrays and a regularized LDA classifier. This model will demonstrate if using information from the entire frequency range and smaller frequency bins improves decoding performance when compared to SF.
- **AE-LDA**: We used the t-f features learnt by AE and a regularized LDA classifier. This provides a baseline to compare SF with AE features.
- **CK-DNN**: We created custom kernels (CK) that act as t-f filters shown in Figure 3b by using an entry of 1 in rows that cover frequency bands from 8-32 Hz and/or 76-100 Hz and 0 for the rest. We have combinations of kernels ranging all temporal scales i.e. 1 to 6 time bins. This give us a total of 63 kernels. We use these custom kernels in the first layer of the DNN described previously for classification. We use this baseline comparison model to be consistent with conventionally used SF while providing the advantage of a DNN for decoding.
- **PCA-LDA**: The weights of an autoencoder with linear activation and lower number of hidden units (n) than number of inputs can span the same subspace as the first n eigen vectors of the covariance matrix of features. Hence, Principal

Component Analysis (PCA) could be a reasonable alternative to the autoencoder based feature extraction methods, specifically in data limited conditions.

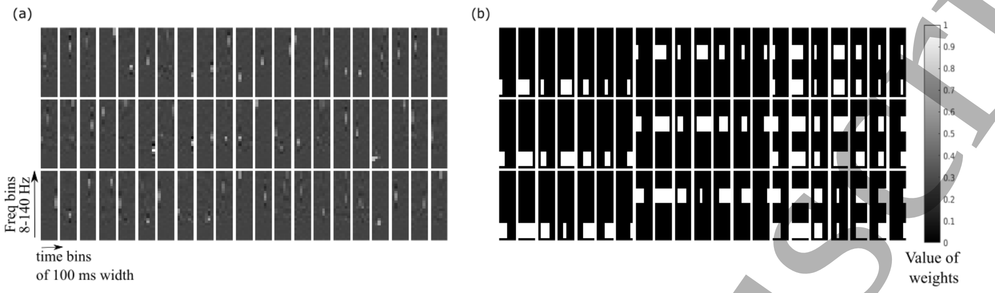

Figure 3: **Comparison of t-f kernels. Kernel ordering in this figure is arbitrary.** Axes are frequency vs time bins as shown in Figure 2a. (a) An example of 63 kernels learnt by autoencoders (b) 63 custom kernels used in CK-DNN.

2.6. Regularization for LDA

We used a regularized LDA where the class covariance is defined by $\Sigma_\gamma = (1-\gamma)\Sigma + \gamma diag(\Sigma)$, where Σ is the empirical, pooled covariance matrix and γ (where $\gamma \in$ [0,1]) is the level of regularization. Optimal values of γ are obtained using Bayesian optimization.

2.7. Robustness of features in noisy conditions

Noise in ECoG recordings due to external sources is a common occurrence. Supplementary Figure 6 shows a sample ECoG signal recorded in the epilepsy monitoring unit outside of controlled experimental conditions. Unlike in controlled experimental settings, brain-computer interfaces have to be robust to external noise when used in real world applications. System design choices, specially for invasive monitoring devices or implants, should also consider the trade-off between battery life and SNR. For example, using low-power amplifiers could result in longer battery life but comes at the expense of SNR. The use of signal processing pipelines robust to noise could play a crucial role in making such design choices.

Due to the use of a non-linear activation, we expect AE based features to be more robust to noise compared to PCA or TF features. To verify this, we simulated frequently occurring spike noise in ECoG by introducing broadband noise of varying levels in randomly selected trials in the training data. We then trained LDA and obtained predictions on test data. Details about noise simulation are in section 5 of Supplementary material.

2.8. Generalization of autoencoder features

We tested our methods on a new dataset to see how well AE features generalize. We used a motor movement and imagery dataset where 5 subjects perform both actual movements of hand and tongue and kinesthetic imagery. During imagery task, subjects were asked to imagine making the movements. However there is no definitive way to verify if they are following the cues during imagery tasks. This dataset was first published in [42]. We used all recorded channels (4̃8-64) for this analysis, not just SMC channels. There are a total of 120 trials for each subject with 30 trials each corresponding to hand and tongue movements and interleaved with rest trials. We used TS learnt using finger flexion dataset to initialize the first layer of network for subjects in this new dataset to test inter dataset transfer learning (ITL-DNN).

2.9. Clustering brain areas

Functional mapping studies use behavioral labels and have some prior constraints on relevant cortical areas when looking for neural correlates. To evaluate if the representations learnt by AE can help identify functionally similar brain areas without any prior knowledge, we clustered electrode channels with the k-means algorithm based on their average activation across finger movement trials. We then validated the clustering results using two metrics.

- **Normalized Mutual Information (NMI)**: We compared our clusters to random clusters to validate that the clusters are informative. We have actual electrode locations (L_{act}) labeled by cortical regions based on anatomical boundaries. These regions do not form a ground truth for clustering, as electrodes from the same regions do not necessarily share the same physiological response patterns. However, physiological responses do tend to be more similar within regions than across regions and thus we expect a "good" clustering to share information with these regions. The random cluster labels (L_{rand}) were obtained by sampling from a uniform distribution of electrode location labels. NMI provides a metric to test if the cluster labels from AE (L_{ae}) are significantly different from L_{rand} and accounts for the varying number of clusters. NMI between two random variables X, Y is calculated using Eq 2, where $I(X, Y)$ denotes the mutual information between X, Y and $H(X)$ denotes the entropy of X. We calculated NMI between our cluster labels and L_{act} and compared this with the NMI between random cluster labels and L_{act}.

$$NMI(X,Y) = \frac{I(X;Y)}{[H(X)+H(Y)]/2} \tag{2}$$

- **Decoding accuracy**: We used groups of channels clustered together to decode finger flexions, using DNNs described earlier, to see if the task relevant electrodes tend to be clustered together.

3. Results

3.1. Decoding finger flexions

Classification performance comparing different feature extraction methods for the six subjects ('A'-'F') is shown in Figure 4. Results for excluded subjects are shown in Supplementary Figure 5. Given the anatomical constraints of movements of ring and pinky fingers, their movements are highly correlated and hence; thus, we merged these two class labels into one class. Models using time-frequency features outperform models using SF, demonstrating the advantage of using features that leverage smaller frequency bins that span the entire broadband frequency range. Statistical significance was evaluated and details are provided in Supplementary Tables 1 & 2. Subject F consistently had low decoding performance across all models; we suspect that this is due to insufficient ECoG grid coverage in SMC regions.

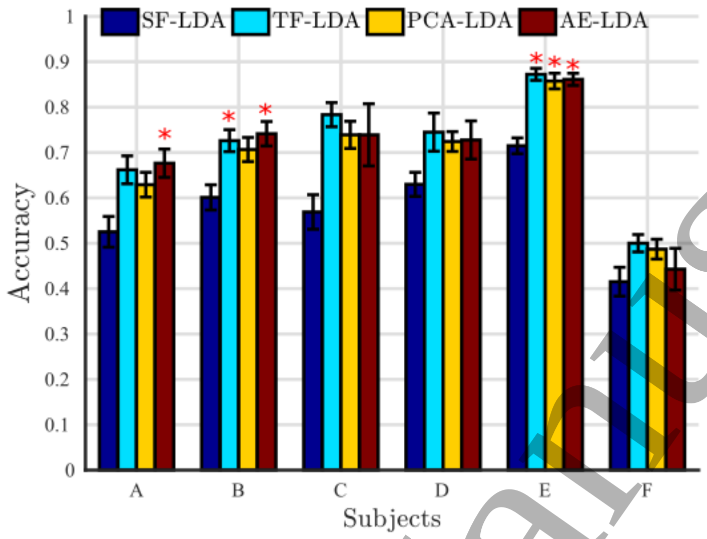

Figure 4: **Comparing different feature extraction methods with LDA classifier.** Mean accuracy with standard error bars over five cross validation folds for 4-class finger flexion classification is shown. Legend: SF-LDA: Spectral Features, TF-LDA: Features are estimated powers in Time-Frequency bins, PCA-LDA: Principal Component projections, AE-LDA: Autoencoder features. Chance level is 28% for random classifier. All the models performed above chance. Models with significantly higher performance compared to SF-LDA are indicated with red asterisks. ANOVA results comparing each condition to the other 3 are in Supplementary Table 1.($P < 0.05$; multiple comparison with one-way ANOVA statistics.)

Although TF-LDA, AE-LDA and PCA-LDA have similar decoding performance against the original datasets collected in this study, autoencoders could provide more robust features under high noise conditions by incorporating a saturating non-linearity and learning sparse representations that are resilient to input dropout. Thus, in Figure 5 we measure the performance of these three decoding strategies with respect to increasing levels of additive spike noise. This analysis demonstrates that AE-LDA is more robust and can be advantages in noisy conditions which are common in EMUs.

LDA uses linear decision boundaries whereas DNN can model complex decision boundaries using non-linear activation functions. We compare how these different classifiers affect decoding performance in Figure 6. We show results from CK-DNN to show performance of spectral features with the advantage of having DNNs. While CK use powers in traditional frequency bands, the TS from autoencoders encode informative patterns learnt from provided data. By using TS from other subjects (TL-DNNs), performance is comparable to that achieved by using features obtained from same subject; this suggests that t-f patterns learnt across subjects hold generalizable information. Given the limited dataset sizes in ECoG studies, using features obtained via TL can be advantageous.

Although the performance of AE-DNN was observed to be on par with that of AE-

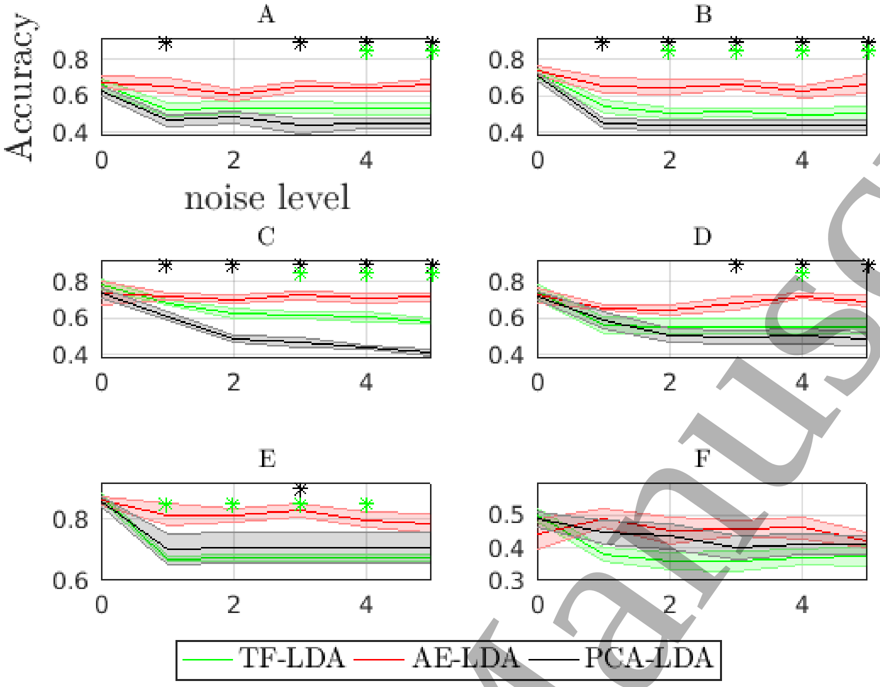

Figure 5: 5 fold cross validation performance (4-class) of TF-LDA, AE-LDA and PCA-LDA with varying noise levels simulating broadband spike noise. Noise levels which have a statistically significant difference between TF-LDA and AE-LDA are denoted with green asterisks and between the PCA-LDA and AE-LDA are denoted with black asterisks. ($P < 0.05$; multiple comparison with one-way ANOVA statistics.)

LDA, DNNs could potentially benefit with increased trial count and have the potential to learn mappings for more complex tasks than LDA. In a more complex task, for example, behavioral labels could be arbitrary relative to the true underlying behavioral intent and structure of neural activity. Conventional techniques like LDA typically assume that there is a prototypical pattern of neural activity for each behavioral label. The DNN, in contrast, has a structure that facilitates a more complex relationship between neural activity and behavioral label. To demonstrate this potential advantage, we consider a scenario in which two classes, thumb and index finger, which have distinct neural activity, are merged to create one label. Neural activity between thumb and index fingers is well distinguishable from one another compared to index and middle fingers. Therefore LDA would be able to find good decision boundaries for classification tasks when thumb, index, and middle fingers are labeled separately. However, when we merge thumb and index classes, effective classification might require a more complex mapping than possible via LDA. Figure 7 shows decoder performance for this modified classification task for AE-LDA and AE-DNN with 50 hidden nodes (arbitrary parameter choice); Supplementary Figure 9 shows the confusion matrices. DNNs fare well when decoding thumb/index vs middle finger demonstrating the comparative advantage DNNs have over LDA when labels are more abstract.

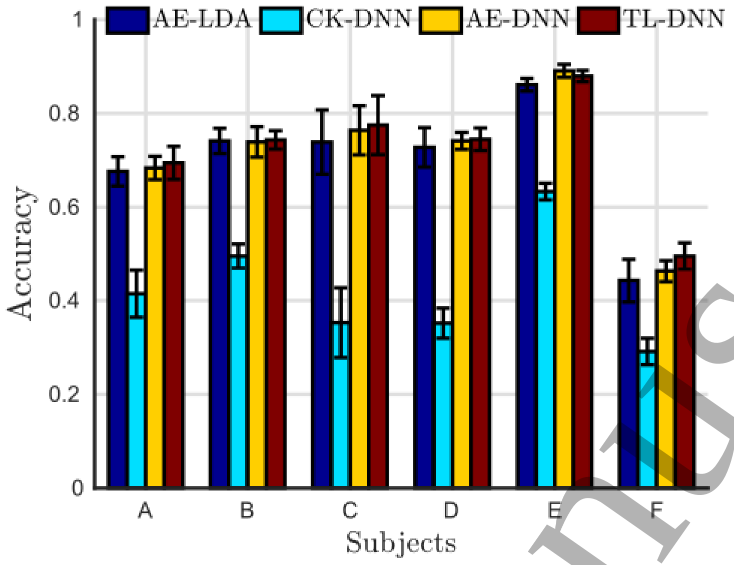

Figure 6: **Comparing different classifiers.** Mean accuracy with standard error bars over five cross validation folds for 4-class finger flexion classification is shown. Legend: AE-LDA: Autoencoder features with LDA classifier, CK-DNN: Custom Kernel features with Deep Neural Network classifier, AE-DNN: Subject specific autoencoder features with DNN, TL-DNN: Transfer Learning features from other subjects with DNN. AE-LDA, AE-DNN, TL-DNN show equivalent decoding performance. Chance level is 28% for random classifier. All the models, except for CK-DNN for subjects C, D and F, performed above chance.

The classification results for hand vs tongue dataset shown in Figure 8 suggest that the decoding method presented in this paper generalizes to other datasets. ITL-DNN performed on par with AE-DNN suggesting the plausibility of inter dataset transfer learning and AE based models outperformed SF-LDA despite the high dimensionality and small dataset size.

3.2. Clustering results

Figure 9 demonstrate that NMI(L_{ae},L_{act}) and NMI(L_{rand},L_{act}) are significantly different (Results from t-test are in Table 3 in Supplementary material). Figure 10a is the brain maps when clustering electrode channels into 4 groups and Figure 10b quantifies classification performance using channels from each cluster. We observed that the cluster that significantly outperforms other clusters has performance comparable to that achieved when using SMC electrodes and tends to include electrodes from dorsal SMC, except for Subject F. Statistical comparisons of performance across clusters are detailed in Table 4 in Supplementary material. This observation is consistent when varying the number of clusters. Further, we performed the same clustering experiment with only SMC electrodes and observed that electrodes in dorsal SMC and ventral SMC are almost always clustered into two groups (see Supplementary Figure 8) and the electrode cluster with higher classification performance is in dorsal SMC. This is

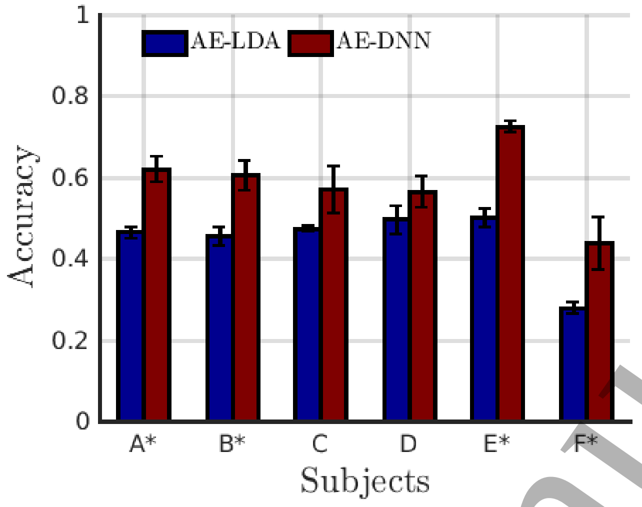

Figure 7: Comparing LDA and DNN when two classes (Thumb,Index fingers) are merged. Mean accuracies with standard error bars for 5 cross-validation folds are shown. Subjects indicated with asterisk have P<0.05 with paired-sample t-Test.

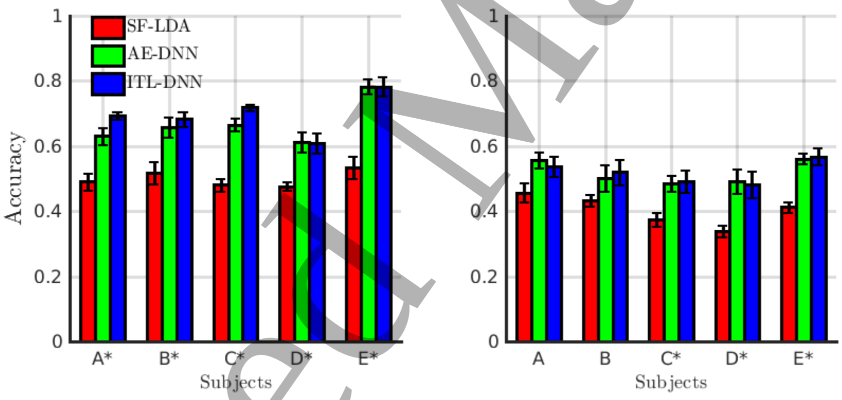

Figure 8: **Decoding tongue, hand and rest trials:** Subjects perform actual movement or kinesthetic imagery of hand and tongue. Mean accuracy with standard error bars over five cross validation folds for 3-class classification is shown. Chance level with random guessing is 37.5%. Actual movement decoding performance (left) is higher compared to kinesthetic imagery (right). For ITL-DNN, template spectrograms trained using finger flexion dataset were used in the first layer of DNN and for AE-DNN, template spectrograms were trained using respective subject's data. Subjects with significant difference ($P < 0.05$; multiple comparison with one-way ANOVA statistics.) between SF-LDA and AE-DNN are indicated with an asterisk.

in agreement with related studies [15, 43] that document distinct dorsal SMC activity during finger flexions. This suggests that the proposed clustering based dimensionality reduction method can be used to identify task relevant channels in an unsupervised fashion.

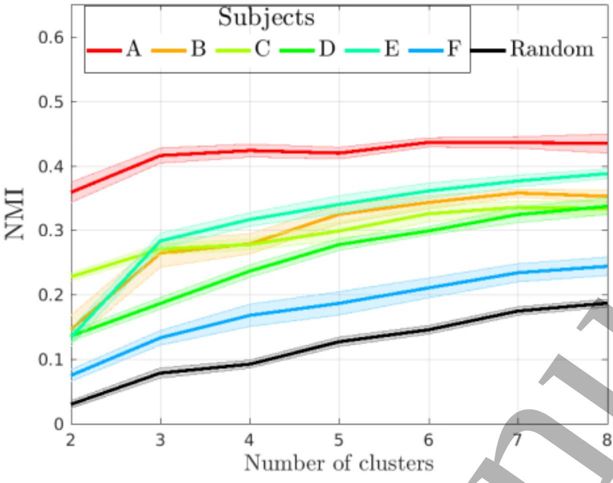

Figure 9: **Comparing Normalized Mutual Information (NMI).** NMI is calculated for 10 random initializations and the mean values with standard error bars are shown.

4. Discussion

During motor activity, significant increases in ECoG signal power occurs mostly in high frequency bands. However, it is difficult to estimate the exact frequency ranges and the timing of increased activity for different electrode channels. We exploit the temporal and frequency information in the data by restructuring signals into a t-f format. If we consider the result from Figure 4, we see that TF-LDA can outperform SF-LDA, suggesting that a finer frequency resolution is advantageous in this task. However, the TF-LDA approach may lead to poor generalization as the number of features grows to the product of number of channels, frequency bins, and time bins. In the current study, the ratio of TF-LDA features to trials was approximately 15 to 1 for most subjects. We noticed that regularization of TF-LDA tended to reduce the effective parameterization of TF-LDA by weighting off-diagonal terms of the covariance matrix in the LDA model close to zero. The resulting model is effectively a Naive Bayes model which assumes all features (i.e. all t-f entries) are independent given the finger being flexed. Thus, we compare these results to PCA-LDA and AE-LDA; AE and PCA (which is the special linear case of AE) can efficiently aggregate data from all frequency bands and time bins based upon underlying regularities (i.e. correlations between t-f entries) in the dataset. AE and PCA allow the models to move away from using fixed frequency band spectral features selected based upon previous literature while controlling the number of features used by the decoding model. The comparable performance for TF-LDA, PCA-LDA, and AE-LDA in Figure 4 suggests that this additional structure can be imposed upon t-f based features without a loss in performance.

Supplementary Figure 7 shows sample t-f responses of an SMC electrode. Since for the majority of electrodes' t-f responses vary in power for specific high frequency

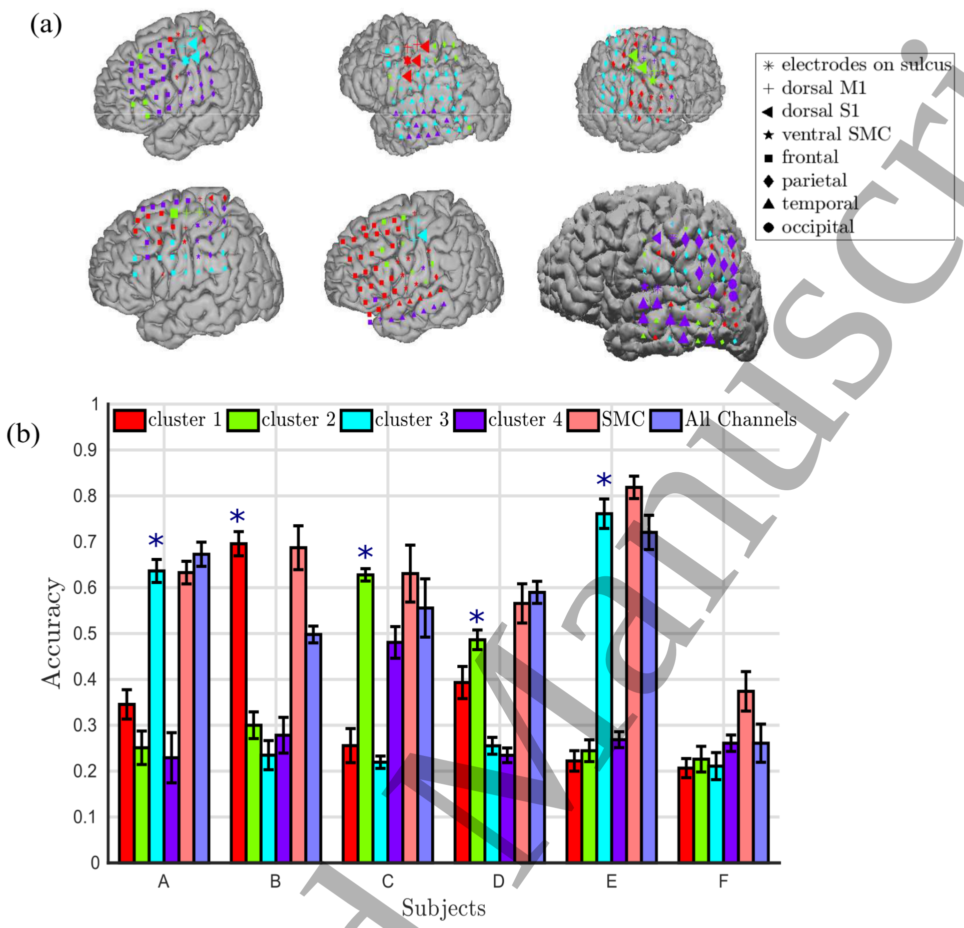

Figure 10: **Functional mapping analysis.** (a): Electrode channels projected onto brain images for five subjects. Colors correspond to different clusters from k-means algorithm. Shapes correspond to original channel locations as obtained from brain images. Note that for subjects A, D and E, electrodes are projected on a standard template and for subjects B, C and F, subjects' original brain maps are used. (b): Decoding performance results using electrode channels from each cluster, SMC channels, and all channels. Colors of cluster bars match with the cluster colors in Figure 10a. The figure shows 4 class classification performance with AE-DNN with fixed (non-optimized) hyperparamter values, $h_1 = 60$ and $h_2 = 10$. Blue asterisks above the bar indicates cluster decoding performance on par with SMC channels decoding performance.

bands in narrow time bins, the autoencoder preserves this information by learning kernels that depict bandpass filters at different time bins as shown in Figure 3a. These template spectrograms generated using AE when used in a convolutional neural network framework serve as filters learnt in a data-driven approach. The AE used in this work uses a non-linear transform, the satlin function. While the addition of this non-linearity does not convey a measurable benefit relative to PCA for the dataset, we demonstrate in Figure 5 how this approach can increase model robustness in higher noise conditions.

DL methods are successfully being employed across various domains such as computer vision, speech recognition, and natural language processing [44, 45, 46]. However, the major impediment for using DL in neurophysiology studies is the lack of sufficient data to train large scale networks with millions of parameters. The proposed DNN architecture has a modest parameter count (of the order of tens of thousands) which can be advantageous for neural decoding when data size is limited. One downside of using DNN methods is the additional effort required for hyperparameter tuning. In our analyses, we used a small grid search on the number of hidden nodes and fixed the learning rate constant. However, one could complete a finer grid search and additional hyperparamater tuning to potentially improve decoder performance at additional computational expense.

Figure 7 demonstrates that DNNs can be advantageous in learning complex mappings under ambiguous or abstract behavioral conditions. This proof-of-concept analysis demonstrates the potential advantage of employing DNN architectures in situations in which behaviors are ambiguous or not well defined. Another advantage of using DNN architectures is that they provide the opportunity to integrate feature extraction methods like AE. By adding the trained AE encoder as an input layer of a DNN classifier, we also have the opportunity to optimize AE weights to better suit the classification problem by backpropagating errors through the single aggregate DNN during training. This could be especially useful in a transfer learning setting, where AE weights are first learnt in an unsupervised fashion and then fine-tuned in combination with DNNs for the classification task at hand.

With advancements in high density ECoG recording technology [47, 48, 49, 50], neurophysiology datasets are likely to see an orders of magnitude increase in data dimensionality. To begin to understand these complex datasets, we must leverage the availability of long duration EMU recordings that can span a few days. However, excluding the recordings collected during structured behavioral tasks, which is of the order of few minutes, a vast portion of the ECoG data is unlabeled or coarsely labeled. The TL-DNN results suggest that a library of t-f patterns learnt across such unlabeled datasets can be used to initiate deep convolutional networks in data limited studies allowing for knowledge transfer.

Surgical procedure for treating epilepsy involves electrically stimulating cortical surfaces through electrodes on grids. With emerging research in high density ECoG, electrode count on grids has increased dramatically. The unsupervised clustering based approach can potentially help clinical functional mapping by reducing the cortical areas of interest to a few localized clusters. While the clustering analysis presented in Figures 9,10 can be done with SF or PCA features, we used the t-f representations from autoencoder, since Figures 4 and 5 suggest that the autoencoder based features are more robust than PCA based features and are more informative than spectral features.

5. Conclusion

Despite the moderate dataset sizes in brain imaging research, deep learning methods are being employed with considerable success in different non-invasive neuro-imaging techniques like functional Magnetic Resonance Imaging (fMRI) and Electroencephalogram (EEG). The standardization of fMRI and EEG data across subjects and experimental studies is also well studied, facilitating the creation of data corpuses and the development of cross-subject analysis techniques. Given the higher relative invasiveness, ECoG datasets are scarce and inter subject variability is high as the electrode grid placement is guided by individual subject's clinical requirements. Recently, the opportunity to record long term ECoG recordings from EMUs along with behavioral data with minimally intrusive video based sensors has facilitated the creation of larger scale data corpus [51, 52]. With the domain knowledge and intuition gained by analyzing small, labeled datasets, we aim to develop generalizable methods for ECoG data analysis that can be applied to analyze bigger, noisier and coarsely labeled datasets. Such methods can enable robust neural decoding for BCIs in complex environments and aid in the development of novel functional brain mapping techniques.

Acknowledgments

We thank the patients and staff at Harborview Hospital in Seattle, WA for their time and dedication. TP is a recipient of IBM Ph.D. Fellowship and this work was further supported by UCSD ECE Department Startup Funding received by VG.

References

[1] Crone NE, Sinai A, Korzeniewska A. High-frequency gamma oscillations and human brain mapping with electrocorticography. Progress in brain research. 2006;159:275–295.

[2] Miller KJ, Shenoy P, Miller JW, Rao RP, Ojemann JG, et al. Real-time functional brain mapping using electrocorticography. Neuroimage. 2007;37(2):504–507.

[3] Leuthardt EC, Schalk G, Wolpaw JR, Ojemann JG, Moran DW. A brain–computer interface using electrocorticographic signals in humans. Journal of neural engineering. 2004;1(2):63.

[4] Crone NE, Miglioretti DL, Gordon B, Sieracki JM, Wilson MT, Uematsu S, et al. Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. I. Alpha and beta event-related desynchronization. Brain. 1998;121(12):2271–2299.

[5] Crone NE, Miglioretti DL, Gordon B, Lesser RP. Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. II. Event-related synchronization in the gamma band. Brain. 1998;121(12):2301–2315.

[6] Bouchard KE, Mesgarani N, Johnson K, Chang EF. Functional organization of human sensorimotor cortex for speech articulation. Nature. 2013;495(7441):327–332.

[7] Pei X, Leuthardt EC, Gaona CM, Brunner P, Wolpaw JR, Schalk G. Spatiotemporal dynamics of electrocorticographic high gamma activity during overt and covert word repetition. Neuroimage. 2011;54(4):2960–2972.

[8] Toro C, Deuschl G, Thatcher R, Sato S, Kufta C, Hallett M. Event-related desynchronization and movement-related cortical potentials on the ECoG and EEG. Electroencephalography and clinical neurophysiology/evoked potentials section. 1994;93(5):380–389.

[9] Kennedy PR, Kirby MT, Moore MM, King B, Mallory A. Computer control using human intracortical local field potentials. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2004;12(3):339–344.

[10] Yanagisawa T, Hirata M, Saitoh Y, Kishima H, Matsushita K, Goto T, et al. Electrocorticographic control of a prosthetic arm in paralyzed patients. Annals of neurology. 2012;71(3):353–361.

[11] Hotson G, McMullen DP, Fifer MS, Johannes MS, Katyal KD, Para MP, et al. Individual finger control of a modular prosthetic limb using high-density electrocorticography in a human subject. Journal of neural engineering. 2016;13(2):026017.

[12] Degenhart AD, Hiremath SV, Yang Y, Foldes S, Collinger JL, Boninger M, et al. Remapping cortical modulation for electrocorticographic brain–computer interfaces: a somatotopy-based approach in individuals with upper-limb paralysis. Journal of neural engineering. 2018;15(2):026021.

[13] Vansteensel MJ, Pels EG, Bleichner MG, Branco MP, Denison T, Freudenburg ZV, et al. Fully implanted brain–computer interface in a locked-in patient with ALS. New England Journal of Medicine. 2016;375(21):2060–2066.

[14] Miller KJ, Leuthardt EC, Schalk G, Rao RP, Anderson NR, Moran DW, et al. Spectral changes in cortical surface potentials during motor movement. The Journal of neuroscience. 2007;27(9):2424–2432.

[15] Miller K, Zanos S, Fetz E, Den Nijs M, Ojemann J. Decoupling the cortical power spectrum reveals real-time representation of individual finger movements in humans. Journal of Neuroscience. 2009;29(10):3132–3137.

[16] Leuthardt EC, Miller KJ, Schalk G, Rao RP, Ojemann JG. Electrocorticography-based brain computer interface-the Seattle experience. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2006;14(2):194–198.

[17] Wander JD, Blakely T, Miller KJ, Weaver KE, Johnson LA, Olson JD, et al. Distributed cortical adaptation during learning of a brain–computer interface task. Proceedings of the National Academy of Sciences. 2013;110(26):10818–10823.

[18] Plis SM, Hjelm DR, Salakhutdinov R, Calhoun VD. Deep learning for neuroimaging: a validation study. arXiv preprint arXiv:13125847. 2013;.

[19] Cecotti H, Graser A. Convolutional neural networks for P300 detection with application to brain-computer interfaces. IEEE transactions on pattern analysis and machine intelligence. 2011;33(3):433–445.

[20] Mirowski P, Madhavan D, LeCun Y, Kuzniecky R. Classification of patterns of EEG synchronization for seizure prediction. Clinical neurophysiology. 2009;120(11):1927–1940.

[21] Wulsin D, Gupta J, Mani R, Blanco J, Litt B. Modeling electroencephalography waveforms with semi-supervised deep belief nets: fast classification and anomaly measurement. Journal of neural engineering. 2011;8(3):036015.

[22] Längkvist M, Karlsson L, Loutfi A. Sleep stage classification using unsupervised feature learning. Advances in Artificial Neural Systems. 2012;2012:5.

[23] Turner J, Page A, Mohsenin T, Oates T. Deep belief networks used on high resolution multichannel electroencephalography data for seizure detection. In: 2014 AAAI Spring Symposium Series; 2014.

[24] Freudenburg ZV, Ramsey NF, Wronkiewicz M, Smart WD, Pless R, Leuthardt EC. Real-time naive learning of neural correlates in ECoG Electrophysiology. International Journal of Machine Learning and Computing. 2011;1(3):269.

[25] Wang Z, Lyu S, Schalk G, Ji Q. Deep Feature Learning Using Target Priors with Applications in ECoG Signal Decoding for BCI. In: IJCAI; 2013.

[26] Masci J, Meier U, Cireşan D, Schmidhuber J. Stacked convolutional auto-encoders for hierarchical feature extraction. In: International Conference on Artificial Neural Networks. Springer; 2011. p. 52–59.

[27] Sanchez J, Principe J, Carmena J, Lebedev MA, Nicolelis M. Simultaneus prediction of four kinematic variables for a brain-machine interface using a single recurrent neural network. In: Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE. vol. 2. IEEE; 2004. p. 5321–5324.

[28] Rao YN, Kim SP, Sanchez JC, Erdogmus D, Principe JC, Carmena JM, et al. Learning mappings in brain machine interfaces with echo state networks. In: Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.. vol. 5. IEEE; 2005. p. v–233.

[29] Gunduz A, Ozturk MC, Sanchez JC, Principe JC. Echo state networks for motor control of human ecog neuroprosthetics. In: 2007 3rd International IEEE/EMBS Conference on Neural Engineering. IEEE; 2007. p. 514–517.

[30] Sussillo D, Nuyujukian P, Fan JM, Kao JC, Stavisky SD, Ryu S, et al. A recurrent neural network for closed-loop intracortical brain–machine interface decoders. Journal of neural engineering. 2012;9(2):026027.

[31] Bashivan P, Rish I, Yeasin M, Codella N. Learning representations from EEG with deep recurrent-convolutional neural networks. ICLR 2016,arXiv preprint arXiv:151106448. 2016;.

[32] Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Cognitive modeling. 1988;5(3):1.

[33] Miller KJ, Hermes D, Honey CJ, Hebb AO, Ramsey NF, Knight RT, et al. Human motor cortical activity is selectively phase-entrained on underlying rhythms. PLoS Comput Biol. 2012;8(9):e1002655.

[34] Miller K, Ojemann J. A library of human electrocorticographic data and analyses. Program No 46908 2016 Neuroscience Meeting Planner San Diego, CA: Society for Neuroscience, 2016 Online;.

[35] Hermes D, Miller KJ, Noordmans HJ, Vansteensel MJ, Ramsey NF. Automated electrocorticographic electrode localization on individually rendered brain surfaces. Journal of neuroscience methods. 2010;185(2):293–298.

[36] DeMers D, Cottrell GW. Non-linear dimensionality reduction. In: Advances in neural information processing systems; 1993. p. 580–587.

[37] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. science. 2006;313(5786):504–507.

[38] Vincent P, Larochelle H, Bengio Y, Manzagol PA. Extracting and composing robust features with denoising autoencoders. In: Proceedings of the 25th international conference on Machine learning. ACM; 2008. p. 1096–1103.

[39] Bengio Y, Courville A, Vincent P. Representation learning: A review and new perspectives. IEEE transactions on pattern analysis and machine intelligence. 2013;35(8):1798–1828.

[40] Vedaldi A, Lenc K. MatConvNet – Convolutional Neural Networks for MATLAB. In: Proceeding of the ACM Int. Conf. on Multimedia; 2015.

[41] Shenoy P, Miller KJ, Ojemann JG, Rao RP. Generalized features for electrocorticographic BCIs. IEEE Transactions on Biomedical Engineering. 2008;55(1):273–280.

[42] Miller KJ, Schalk G, Fetz EE, Den Nijs M, Ojemann JG, Rao RP. Cortical activity during motor execution, motor imagery, and imagery-based online feedback. Proceedings of the National Academy of Sciences. 2010;107(9):4430–4435.

[43] Kubanek K Miller, Ojemann J, Wolpaw J, Schalk G. Decoding flexion of individual fingers using electrocorticographic signals in humans. Journal of neural engineering. 2009;6(6):066001.

[44] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems; 2012. p. 1097–1105.

[45] Hinton G, Deng L, Yu D, Dahl GE, Mohamed Ar, Jaitly N, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal Processing Magazine. 2012;29(6):82–97.

[46] Collobert R, Weston J. A unified architecture for natural language processing: Deep neural networks with multitask learning. In: Proceedings of the 25th international conference on Machine learning. ACM; 2008. p. 160–167.

[47] Viventi J, Kim DH, Vigeland L, Frechette ES, Blanco JA, Kim YS, et al. Flexible, foldable, actively multiplexed, high-density electrode array for mapping brain activity in vivo. Nature neuroscience. 2011;14(12):1599–1605.

[48] Chang EF. Towards large-scale, human-based, mesoscopic neurotechnologies. Neuron. 2015;86(1):68–78.

[49] Hermiz J, Rogers N, Kaestner E, Ganji M, Cleary DR, Carter BS, et al. Sub-millimeter ECoG pitch in human enables higher fidelity cognitive neural state estimation. NeuroImage. 2018;176:454–464.

[50] Ganji M, Kaestner E, Hermiz J, Rogers N, Tanaka A, Cleary D, et al. Development and Translation of PEDOT: PSS Microelectrodes for Intraoperative Monitoring. Advanced Functional Materials. 2018;28(12):1700232.

[51] Gabriel P, Doyle WK, Devinsky O, Friedman D, Thesen T, Gilja V. Neural correlates to automatic behavior estimations from RGB-D video in epilepsy unit. In: Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the. IEEE; 2016. p. 3402–3405.

[52] Wang XRN, Farhadi A, Rao R, Brunton B. AJILE Movement Prediction: Multimodal Deep Learning for Natural Human Neural Recordings and Video. In Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI). 2018;.