UNIVERSITY OF CALIFORNIA, SAN DIEGO

Towards Practical Neural Prosthetic Interfaces

A Thesis submitted in partial satisfaction of the
requirements for the degree
Master of Science

in

Electrical Engineering
(Intelligent Systems, Robotics, and Control)

by

Aashish N. Patel

Committee in charge:

Professor Vikash Gilja, Chair
Professor Tzyy-Ping Jung
Professor Kenneth Kreutz-Delgado
Professor Terrence J. Sejnowski

2017

The Thesis of Aashish N. Patel is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2017

Simplicity is the ultimate sophistication

—Leonardo da Vinci

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I have had the privilege of being a part of a brilliant research family. Vikash, it has been a pleasure learning from you over the past few years. You have been a great mentor who has constantly motivated me to be a better researcher. Without your guidance and steadfast patience, it would not have been possible to accomplish even a fraction of what I did.

It has been a trip, TNEL. Venkatesh, we accomplished great things through the countless sleepless nights. It would not have been as much fun without you. Tejaswy P. and Paolo G., thank you for immeasurable help and for keeping things interesting - there was never a dull moment. To my research family at SCCN – Tzyy-Ping and Siddharth, it is an understatement to say it has been a great joy working with you both over the past two years.

I have been even more privileged to have loving and supportive friends and family. Mom and Dad, thank you for believing in me and teaching me to always keep my goals in sight. Feena, your support got me through tough times. You continue to push me to aim for the stars, but reach further. Benjamin M. and Kevin T., thank you for always being available to lend a helping hand. And to my countless friends and family who have supported me through my endeavors, I could not be more grateful.

Chapters 3, 4, and 5 in this thesis are adopted from work conducted with Venkatesh Elango, Siddharth, Dr. Kai J. Miller, Dr. Terrence J. Sejnowski, Dr. Tzyy-Ping Jung, and Dr. Vikash Gilja. Contributing authors are acknowledged at the end of each respective chapter.

ABSTRACT OF THE THESIS

Towards Practical Neural Prosthetic Interfaces

by

Aashish N. Patel

Master of Science in Electrical Engineering
(Intelligent Systems, Robotics, and Control)

University of California, San Diego, 2017

Professor Vikash Gilja, Chair

The connection between our brain and our muscles via the peripheral nerves enables us to communicate and act upon the world. Individuals suffering from disease and injury that affect these connections have limited treatment options and often rely upon prostheses. A fundamental challenge for such prostheses is the development of a control system that is capable of interpreting desired intentions accurately. Existing control schemes utilize activation by other functional systems using clever mechanical linkages or electromyography signal based control. Increasingly, neural interfaces are being considered for these applications due to their

potential ability to restore behavioral function without co-opting existing motor functions. Implanted electrode based neural interfaces, such as electrocorticography (ECoG) hold promise for providing high signal to noise ratio measurements that are stable over long time periods while electroencephalography (EEG) provides convenient, non-invasive scalp surface measurements. Although these techniques provide powerful approaches to measuring neural signals, their application currently yields limited performance in neural interfacing applications.

Addressing the complexity of interpreting neural signals, we explore different neural decoding algorithms and evaluate their performance and limitations. Taking advantage of the time-varying properties of neural oscillations, we propose algorithms for sequence learning and sequence transfer learning. We also explore the different user interaction scenarios and the associated performance limitations. Furthermore, in advancing the adoption and understanding of non-invasive neural interfaces, we outline a portable system capable of measuring complex montages of multi-modal bio-sensing data. The perspective gained from these analyses provide better insights to practical neural interface design.

Chapter 1

Introduction

1.1 Motivation

Our ability to interact with the world through communication and movement are fundamental capabilities that are difficult to live without. Unfortunately, millions of people around the world live with neurodegenerative disorders like amyotrophic lateral sclerosis (ALS) and multiple sclerosis (MS) or suffer from spinal cord injuries leading to partial and full paralysis. These individuals have varying ranges of obstructed interaction with the most extreme conditions resulting in locked-in syndrome where the person can no longer communicate with others.

Numerous avenues of research are devoted to developing approaches for limiting and ultimately reversing neuron damage from diseases or physical trauma. Other approaches have focused on improving the quality of life for individuals through either pain or symptom mediation. While some of these techniques will ultimately eliminate the need for any prosthetic device, particularly for nerve damage cases, the development of neural prosthetic interfaces can address interaction capabilities for individuals with irreparable nerve or limb damage.

1.2 Existing Solutions

There are many prostheses available to individuals suffering from neuron damage. These devices are typically well-designed mechanical systems that benefit from under-actuated control schemes as they rely on functioning movement capabilities. Examples of such devices can be found mostly for locomotion and hand prostheses: Jaipur foot [1, 2], Deka arm [3], etc. While providing partial restoration of function, however, there are imposed hardships on other body components, leading to the belief that more naturalistic prostheses and control schemes are better for usability and health [4].

Although more advanced motorized prostheses can be controlled by other functional movement capabilities, alternative control signal sources are currently under consideration. Electromyography (EMG) is a promising technique that uses the neuron activation in the peripheral nervous system to activate the prostheses [5]. While this works for partial amputations, it has limitations in the number of degrees of freedom capable of being simultaneously controlled and instead relies on clever control schemes for driving complex prosthetic actuators [6].

Neural interfaces that interact directly with the brain are also being evaluated for use in controlling prostheses. The non-invasive solution involves electroencephalography (EEG) where the measurements are made on the surface of the scalp. This technique, due to the spatial averaging that occurs from neural signals traversing the skull and scalp, has limited resolution and typically utilizes stereotypical responses such as SSVEP, RSVP, or P300 responses for control [7, 8, 9]. Implanted neural interfaces, such as electrocorticography (ECoG) and depth electrodes, are also under consideration as control signal sources [10, 11, 12, 13]. These techniques yield higher signal quality and provide the ability to measure neural activity at greater resolution than typical EEG approaches. A challenge to ECoG

adoption is the required surgery to implant the electrodes, and the need for long-term electrode measurement stability [14]. However, promising devices like the NeuroPace RNS are tackling this problem with implantable long-term, closed-loop measurement and stimulation systems. Finally, optogenetics interfaces are being explored for their efficacy due to its ability to measure from a large quantity of individual neurons simultaneously [15]. The introduction of viral vectors to genetically modify neurons to allow for fluorescent behavior limits the near-term usability of such systems.

1.3 Challenges

Considering the solutions and associated limitations of existing prosthetic interfaces discussed in the previous section, the focus of this thesis is on neural interfaces that directly measure neural activity from the brain. Addressing ECoG measurements, numerous challenges must still be overcome before utilization as a long-term neural prosthetic interface. Primarily, the neural decoders must improve accuracy for a wide range of behaviors and symbol counts. It is important to expand this capability to allow for higher rates of communication and more natural interactions [16].

While decoder accuracy can improve the quality of the prosthetic, it has little utility if the system is overlooked for ones with more usable interfaces. It is clear that with consideration of user-interface design, neural prosthetic systems can be integrated into systems that are comfortable, and even enjoyable for users [5].

Moreover, a major challenge is being able to translate and integrate research into products for individuals. Typical experiments involving ECoG recordings must occur in the hospital due to portability constraints. EEG systems that are

portable and easy to interface with, however, exist. The experiments, however, are typically constrained to laboratory settings in stationary positions or controlled movement scenarios due to the inability to integrate different sensors in a portable system while retaining data integrity and synchronization [17].

1.4 Chapter Structure

This thesis is composed of five core chapters in addition to the Introduction and Conclusion.

In Chapter 2, we provide an overview of different neural decoding algorithms currently used in ECoG brain-computer interfaces (BCIs). The state-of-the-art decoders are evaluated for the dataset utilized in the remainder of the thesis. The resulting analyses provide clear insight to performance trade-offs of each algorithm and provides motivation for sequence decoding.

In Chapter 3, we analyze existing sequence linear discriminant and hidden Markov model approaches and propose using long short-term memory networks. Furthermore, a transfer learning architecture is proposed with results demonstrating robust decoder performance that exceeds performance of existing classifiers.

We outline a series of user-interaction schemes in Chapter 4 and utilize sequence decoding with hidden Markov models to elucidate associated advantages and limitations. Additionally, we propose alternative designs to relaxing timing rigidity in existing control schemes.

Lastly, we propose a system developed to help advance translatable research in Chapter 5. The device is capable of collecting EEG data and other bio-sensing data simultaneously. Moreover, the system provides an untethered platform for conducting complex experiments in real-world environments. Utiliz-

ing the proposed device and software components, we provide a framework for better understanding neural correlates to behavior and emotion.

Chapter 2

Neural Decoder Models

Neural decoder models can be divided into two classes, linear and sequence models. In this chapter, the dataset utilized for both models is outlined with emphasis placed on nuances between the trialization techniques. Furthermore, the linear models and corresponding results are discussed in this chapter with sequence model approaches being presented in Chapter 3.

2.1 Data Description

Neural signals were recorded from nine patients being treated for medically-refractory epilepsy using standard sub-dural clinical electrocorticography (ECoG) grids. Each array (Ad-Tech, Racine, WI) contained 34 - 64 platinum electrodes. Each electrode had a 4 mm diameter (2.3 mm exposed), 1 cm inter-electrode distance, and were embedded in silastic. The experiment was a finger flexion task where subjects wearing a DataGlove (Fifth Dimension Technologies, Irvine, CA) were asked to flex a finger for two seconds based on a visual cue [18]. 30 cues per finger were collectred across the 10 minutes of experimentation with and average of 3 - 5 flexions measured per cue segment. All subjects participated in a purely

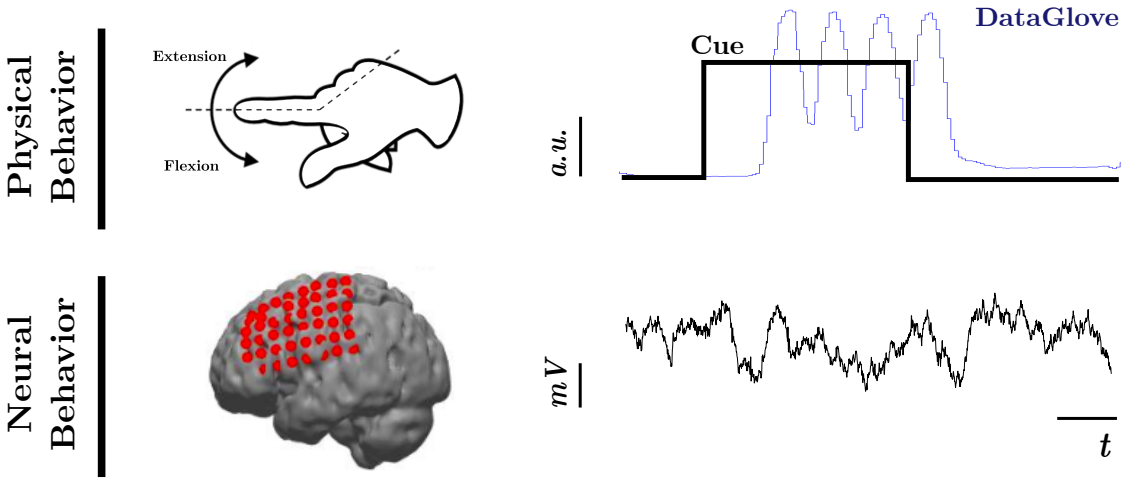

Figure 2.1: Overview of finger flexion experiment. The top row depicts the physical behavior measured during during the cue stimuli. The bottom row depicts a sample subject brain with implanted ECoG grid and measured neural trace.

voluntary manner, after providing informed written consent, under experimental protocols approved by the Institutional Review Board of the University of Washington. All patient data was anonymized according to IRB protocol, in accordance with HIPAA mandate. This data has been released publicly [19]. Figure 2.1 depicts the experimental structure.

2.2 Signal Conditioning

Different subject selection criteria were utilized for fixed length and variable length sequence decoders. For the fixed length sequences (linear models), three subjects were excluded from the analysis due to mismatches in behavioral measurement and cue markers. For the variable length sequences (sequence models) discussed in Chapter 3, two additional subjects were excluded due to insufficient coverage of the sensorimotor regions.

Analyzing the subject specific neural data, electrodes are rejected if the channel data exceed two standard deviations of the mean signal. The remaining electrodes are conditioned to eliminate line noise, and then instantaneous spectral power in the high band range (70 - 150 Hz) for each electrode is extracted for use as the features for the classifier [20]. The data is segmented using only cue timing information resulting in 27 - 29 trials per finger (5 classes), with each trial having a $\sim$ 2100 ms duration. Results using region filtering contain 6 - 8 electrodes for each subject and cover the sensorimotor region due to its importance in motor planning. Where noted, frontal cortex electrodes are also evaluated due to their involvement in higher-order processing for motor tasks.

2.2.1 Linear Models

For the linear models analyzed, a strong criteria for data binning was utilized to maximize decoder performance. Specifically, the DataGlove recordings are used for segmentation opposed to the cue information. To convert the continuous behavioral DataGlove signal to a discrete activation, the data was first coarsely thresholded using 40% above the detrended baseline amplitude relative to the maximum value. Finger movement amplitudes were binned to find the baseline references. The center of the heaviest bin was set as the baseline. The true behavioral onset time was then found within a window preceding the coarse threshold marker between $t - 100$ ms and t. Using only the first flexion time marker in a cue, the data is labeled as thumb, index, middle, ring, or pinky. Trialization of the neural data was computed using the average high band power for each electrode for non-overlapping windows of 750 ms beginning at the behavior. Window and offset parameters were selected from the results of a grid search.

2.2.2 Sequence Models

Given the nature of the decoding problem presented, sequence models are a sensible approach. While not explored in this chapter, a detailed analysis is provided in Chapter 3. The sequence models utilize a slightly different alignment scheme than the one proposed for the linear models. The cue data, for one, is the only alignment information that is used for data trialization. Additionally, a smaller window of 150 ms is used for computing the average high band power for each electrode. This allows for better resolution of the neural oscillations compared to the larger averaged response used in the linear model analysis.

2.3 Linear Models

The standard formulation of linear models are characterized for use in individual finger classification: Naive Bayes (NB), Linear Discriminant Analysis (LDA), Support Vector Machine (SVM) with linear kernel, and Logistic Regression (LR). The features used are a *single* time bin from each flexion sequence per electrode. All results were five-fold cross-validated. As shown in Figure 2.2A, the movement and non-movement states can be classified perfectly using LDA. Using this as a basis for providing only movement data to each of the linear classifiers, a hierarchical classification structure is utilized. We observe that in this setup that LDA, LR, and SVM all perform fairly well with LR performing the best across all subjects.

2.4 Results

The 5-class discrimination problem results are shown in Figure 2.3A. The best models perform below 70% on average across patients, but have a large variance. This inter-patient spread is likely a function of the behavior-cue coherence. Exploring the class discriminability, confusion matrices for the movement and finger detectors are shown in Figure 2.2. The movement detector performs exceptionally with near perfect classification across all patients. This is a result of the movement task mapping to an amplitude modulation problem for regions of interest. While only providing binary output, this provides a basis for simple decoders that are capable of very accurately providing feedback. Examining the finger decoding, it is clear that the primary three fingers are capable of decoding at a much greater accuracy than the ring and pinky fingers. Reviewing the DataGlove, it is clear that there is a significant amount of noise present during these movements. This can be explained by the limited independent articulation capabilities of the two fingers - not only is there less cortical area for motor planning, but paired movement causes afferent innervation leading to difficulty in disambiguation. In Chapter 5, symbol count trade-offs are analyzed to evaluate the utility of this decoding.

Reviewing the linear classification performance using only a single, large window per electrode, it is clear that the average shift in amplitude across the different electrodes provides very useful information that can be used for a decoder. In particular, simple movement detectors that are able to perfectly separate classes can provide reliable binary output. Limitations with the ability to accurately decode noisy classes requires the exploration of better trialization and decoding strategies.

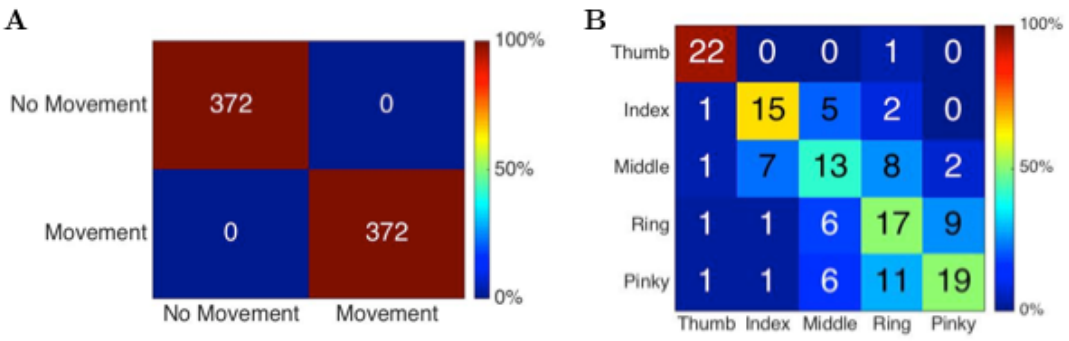

Figure 2.2: Linear discriminant analysis classification confusion matrices of movement-rest and 5-finger flexion.

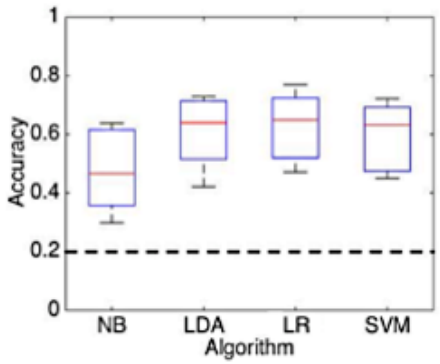

Figure 2.3: Decoding classification accuracy for linear models. Naive Bayes (NB), linear discriminant analysis (LDA), logistic regression (LR), and support vector machines with linear kernel (SVM) result statistics are shown across all 6 subjects.

Chapter 3

Decoding Neural Sequences

3.1 Sequence Linear Discriminant Analysis

The standard LDA formulation evaluated in the previous chapter considers each sample in time. While this representation is powerful when the useful features are contained in the averaged window, it is difficult to capture behaviors that have a time scale larger than the window. To allow for temporal modeling, the input space needs to be modified to allow use of the time bins as features (LDA-Seq) [21]. Therefore the decoder is constructed to take in power from different frequency bands over a time period of interest per electrode as features. LDA-Seq is supplied the feature vector, $\mathbf{x}$, and the class label, y, to decode the most likely class label, $\hat{y}$, by

$$\mathbf{x}|y \sim \mathcal{N}(\mu_y, \mathbf{\Sigma}_y)$$

$$\hat{y} = \arg\max_y P(\mathbf{x}|y)$$

This maximum likelihood formulation can be extended to any generative model by the changing the distribution of $P(\mathbf{x}|y)$.

We see in Results that this approach yields high accuracy. As such, we emphasize that LDA-Seq results in a strong baseline performance because it explicitly models the temporal response as a function of neural activity relative to movement onset. This is in contrast to hidden Markov models and long short memory models discussed later for which learned parameters do not change as a function of time relative to movement onset, but instead, temporal dynamics are captured by parameters that are constant across time. Consequently, LDA-Seq works well for structured experimental tasks, but does not generalize well to unstructured realistic scenarios and may be prone to over-fitting as electrode count and trial duration increase.

3.2 Hidden Markov Models

Hidden Markov models provide a probabilistic model framework that can model sequential data. At every time step in an N-state HMM, an observation can be obtained from the emission probability distribution conditioned on the latent state. As we are using continuous surface local field potential, we can model the emission probability as a Gaussian distribution [22, 23]. The latent state dynamics follow the Markov property and can be described by a probability transition matrix. Expectation Maximization (EM) is used to learn the HMM parameters. An HMM is trained for each class, and classification is performed by selecting the class that maximizes the sequence likelihood.

We construct an HMM baseline model with 1-state and 2-states using Gaussian emissions to explore if the finger flexion can be modeled with more expressive

temporal models. The 2-state HMM is a standard ergodic HMM [24] allowing transitions between both the states. The 1-state HMM, however, is a special case as it does not have any state dynamics and makes an independence assumption for samples across time. The 1-state HMM is therefore specified by

$$P(O^k|y) = P(o_1^k|y)P(o_2^k|y)\dots P(o_t^k|y) \tag{3.1}$$

$$o_t|y \sim \mathcal{N}(\mu_y, \mathbf{\Sigma}_y) \tag{3.2}$$

where O^k denotes the observed sequential data and y denotes the class. The full derivation of Expectation-Maximization (EM) and Viterbi decoding for the Gaussian emission HMM are found in Appendix B.

One of the advantages of HMMs are the small amount of parameters that need to be learned. And utilizing a Gaussian Mixture Model fit of the training data for initialization, convergence occurs in few iterations. The parameter count is computed as a function of features, f, states, n, and classes, c

$$params = \left(f \times \frac{f+1}{2} + f + n \times (n-1)\right) \times c$$

One-state and two-state models are explored in detail and presented in Results.

3.3 Long Short-Term Memory Networks

Long Short-Term Memory (LSTM) networks have demonstrated an ability to integrate information across varying time-scales and have been used in sequence modeling problems including speech recognition and machine translation [25, 26, 27]. While previous work using LSTMs has modeled time varying signals, there has been little focus on applying them to neural signals [28].

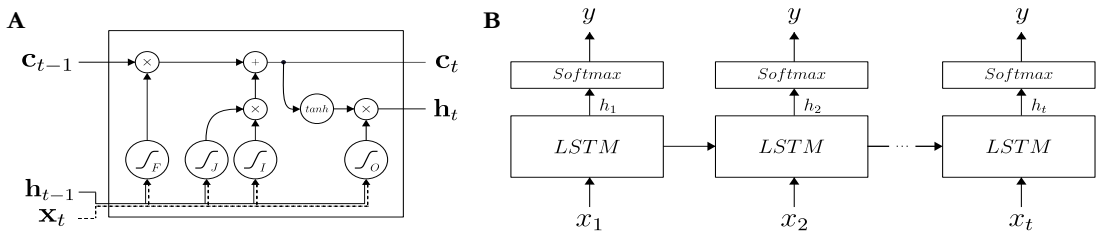

Figure 3.1: A schematic depicting the Long Short-Term Memory (LSTM) model architecture. (A) Gating and activation functions for a single unit. (B) Unrolled LSTM network with Softmax output at every time step during training.

The single recurrent LSTM cell proposed by [29] is utilized for neural decoding and is shown in Figure 3.1A. The model equations are fully specified in Appendix A. At every time step during training, the label is provided to allow error propagation during intermediate steps rather than only after the full sequence has been evaluated [30] and is depicted in Figure 3.1B. Additionally, different initialization schemes, dropout, regularization, complexity, and training strategies were explored. The parameters were fit to a single patient and the model applied to all subjects. Prediction accuracy was evaluated at the last time step of the sequence. In contrast to HMMs, a larger number of parameters need to be learned. The amount can be computed as a function of features, f, hidden units, h, and classes, c:

$$params = h \times (h + f + 1) \times 4 + (h + 1) \times c$$

3.3.1 Initialization Schemes

Three initialization schemes were first explored: random, language model, and sequence autoencoder [31]. The random initialization established the naive

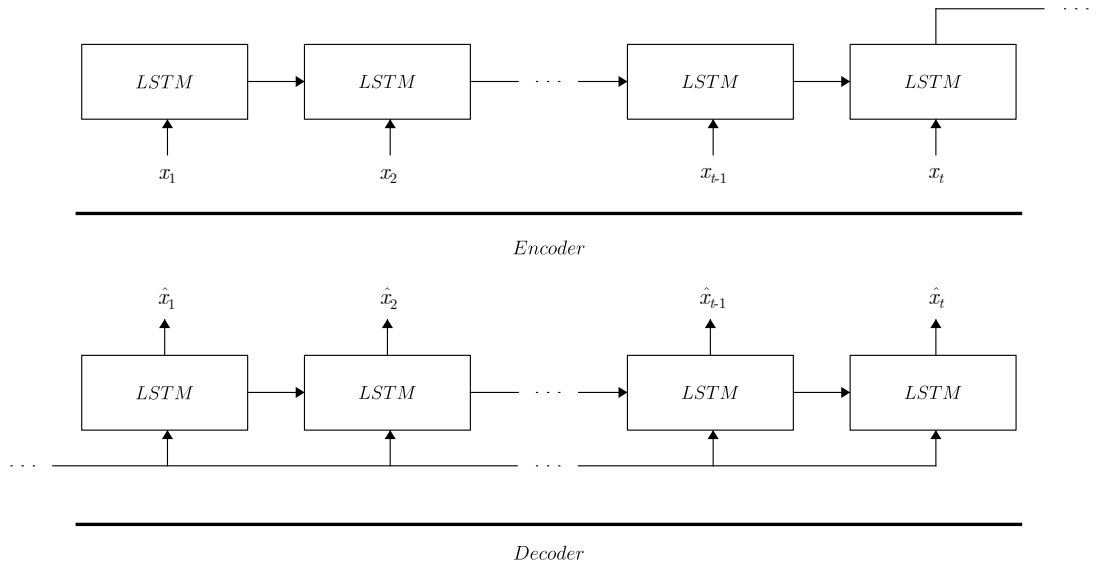

Figure 3.2: Modified sequence autoencoder architecture used for initialization of LSTM network weights.

approach to initializing the model where the latter two approaches explored more structured approaches. After evaluating the different schemes and the associated performances, we ultimately decided to build on the sequence autoencoder and utilize a modified version where the encoder and decoder weights are not shared (Figure 3.2). This approach is similar to the LSTM autoencoder from [32].

3.3.2 Dropout and Regularization

Due to the limited number of samples utilized for training the model, different dropout and regularization parameters were searched to prevent over-fitting. Regularization was ultimately not used due to lack of performance improvement. Dropout on the recurrent states [33] with a probability of 0.3 was used, however. Dropout was not utilized on the input layer due to the limited dimensionality.

3.3.3 Network Complexity

To control the model complexity, different internal recurrent state activation functions and hidden unit counts were evaluated. Exploring the recurrent state architecture, gated recurrent units [27] were utilized, but to a slight performance decrease. Exploring a range of network complexities, it was determined that 100 hidden units provided the best performance for the training time. We found that it is possible to achieve the same performance using 10 hidden units as well, but at a much greater training epoch. Moreover, exceeding 100 hidden units provided no measurable improvement in performance.

Establishing the network complexity, approaches for improving network training performance were explored. In particular, two learning strategies were identified: full-sequence learning and curriculum learning. The standard approach uses the full-sequence during training across all epochs. Curriculum learning, originally proposed by [34], increases the training sample difficulty across epochs and [35] proposed adding a probability for selecting the full-sequence or the curriculum-sequence. This approach is especially useful in situations where optimization using stochastic gradient descent is difficult [35]. In our task, difficulty is directly related to the length of the sequence. Defining L as the full-sequence length, the curriculum-sequence length, l, is scaled by the current epoch, k, using $l = \min\{k/8 + 3, L\}$. Each sample in a batch is drawn from a uniform distribution to be of length L or l.

Lastly, to prevent overfitting, training was stopped at 75 epochs for all evaluations and subjects. All hyper-parameters were selected through optimization on a single patient, Subject B, and used for all subjects. Models were trained through backpropagation using Adam [36].

3.4 Transfer Learning

We next explored the possibility of transferring the representation learned by the LSTM from a subject, S_1, onto a new subject, S_2. Typical transfer learning approaches keep the lower-level layers fixed, but re-train the higher-level layers [32, 37]. Due to the unique electrode coverage and count as well as physiological variability across subjects, this approach yielded poor results. Accounting for these factors, we propose using an affine transformation to project the data from S_2 onto the input of an LSTM trained on S_1 as we might expect a similar *mixture* of underlying neural dynamics across subjects. The fully connected affine transformation is specified by

$$x_t^{S_2'} = W_{xx} x_t^{S_2} + b_x$$

where W_{xx} and b_x are the weights and biases of the affine mapping, $x_t^{S_2}$ and $x_t^{S_2'}$ are the original and the transformed sequential data from S_2 respectively.

Using hyper-parameters outlined in the single subject LSTM model, a two-step training process, shown in Figure 3.3A, is utilized. The first step trains the LSTM using all of the data from S_1. Upon fixing the learned weights, the fully connected affine layer is attached to the inputs of the LSTM and trained on S_2 data. To establish that the affine transformation is only learning an input mapping and not representing the neural dynamics, a baseline comparison is utilized where the Step 1 LSTM is fixed to a random initialization and only the $Softmax^{S_1}$ is trained, this is shown in Figure 3.3B. A more detailed explanation of the LSTM TL architecture can be found in [38].

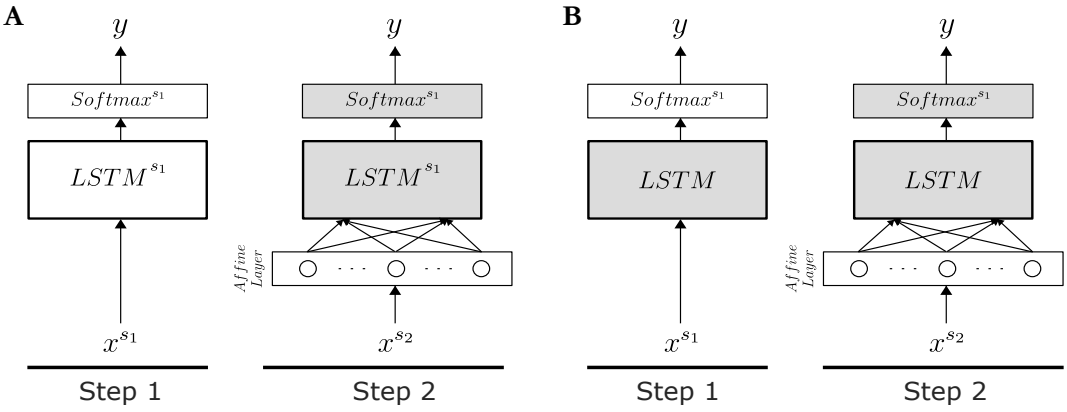

Figure 3.3: A schematic depicting the LSTM transfer learning model architecture. Where gray indicates fixed and white indicates learned weights, both (A) and (B) depict a 2-step training process with Subject 1 (S_1) training in step 1 and Subject 2 (S_2) training of only the affine layer in step 2. (A) Transfer learning model (TL) training the LSTM and Softmax layers for S_1 in step 1. (B) Randomly initialized LSTM layer and training only the Softmax layer for S_1 in step 1.

Table 3.1: Summary comparing the average accuracy of the LSTM model with existing approaches after 20 random partitions of all the shuffled training data.

Model	Subject A	Subject B	Subject C	Subject D
LDA	0.50	0.53	**0.79**	0.64
HMM - 1s	0.51	0.61	0.69	0.65
HMM - 2s	**0.53**	0.59	0.68	0.60
LSTM	0.51	**0.62**	0.75	**0.69**

3.5 Results

3.5.1 Sequence Models

We found that the HMM approach is capable of achieving performance of LDA-Seq while relaxing the timing constraints imposed by the latter. Furthermore, we found that LSTMs trained on an individual subject perform comparable to state-of-the-art models with sufficient training samples.

First we establish the performance of the baseline models (Table 3.1). In-

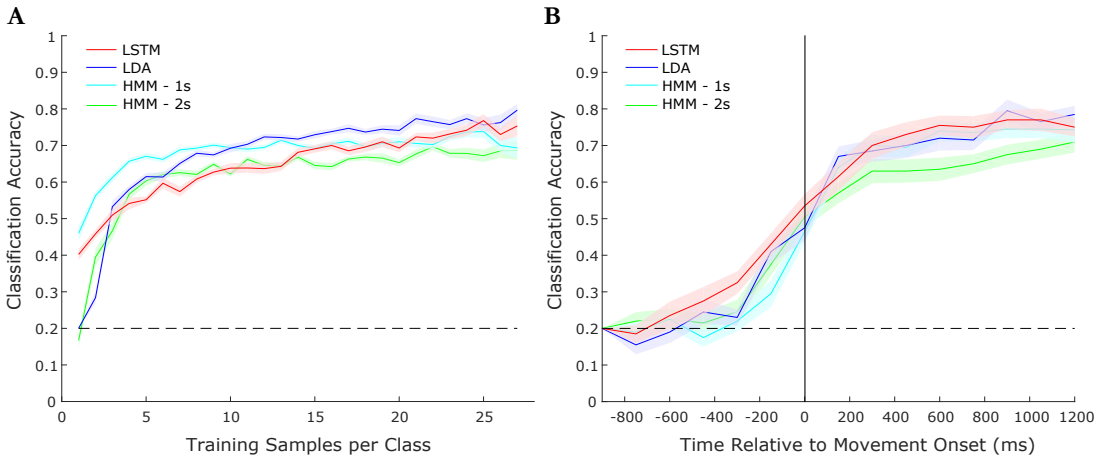

Figure 3.4: Model performance comparison for sequence learning techniques. (A) Accuracy as a function of the amount of training samples and (B) Accuracy as a function of time with respect to movement onset evaluated for different models and using all available training data. Results from representative subject C are shown. Error bars show standard error of mean using 20 random partitions of the shuffled data.

terestingly, we observe that increasing the complexity of the HMM marks little improvement in the classification accuracy and typically results in decreased accuracy at low sample counts due to the increased complexity. Additionally, while LDA performs comparably for three subjects, it performs much better than the other models for Subject C due to the increased alignment between cue and observed behavior. This is expected because, as noted, the LDA formulation is better suited to take advantage of the time alignment in the experiment.

Examining the performance of the LSTMs (Table 3.1), we demonstrate that the proposed model is better able to extract information from the temporal variability in the signals than HMMs and is able to achieve performance comparable to the best baseline for each of the subjects. Consequently, we observe across most subjects LSTMs are able to exceed performance of both HMM models and LDA. Even for Subject C, the LSTM model is comparable to LDA without making the

temporal alignment assumption.

Further investigating model time dependence, accuracy is evaluated using neural activity preceding and following behavioral onset. As LDA explicitly models time, a series of models for each possible sequence length are constructed. Depicted in Figure 3.4B, we observe that the LSTM is slightly better able to predict the behavioral class at earlier times compared to HMMs and is comparable to LDA across all times.

3.5.2 Transfer Learning

Historically, neural prostheses must be tuned frequently for individual subjects to account for neural variability [39, 40]. Establishing LSTMs as a suitable model for decoding neural signals, we explored their ability to learn more robust, generalized representations that can be utilized across subjects.

We demonstrate that learning the affine transformation for the input, it is possible to relax the constraint of knowing exactly where the electrodes are located without having to retrain the entire network. First examining the baseline condition in order to assess the ability for the affine layer to learn the underlying neural dynamics, the S_1 LSTM weights were randomly initialized and fixed as outlined in Figure 3.3B. Looking at Figure 3.5A, the model performs slightly above chance, but clearly is unable to predict behavior from the data. The TL model where only the affine layer is trained in Step 2 for S_2 (Figure 3.5A) performs comparably to the subject specific model for S_2. The notable advantage provided by TL is that there is an increase in loss stability over epochs, which indicates a robustness to overfitting. Finally, relaxing the fixed $LSTM^{S_1}$ and $Softmax^{S_1}$ constraints, we demonstrate that the TL-Finetuned model achieves significantly better accuracy than both the best linear model and subject specific LSTM.

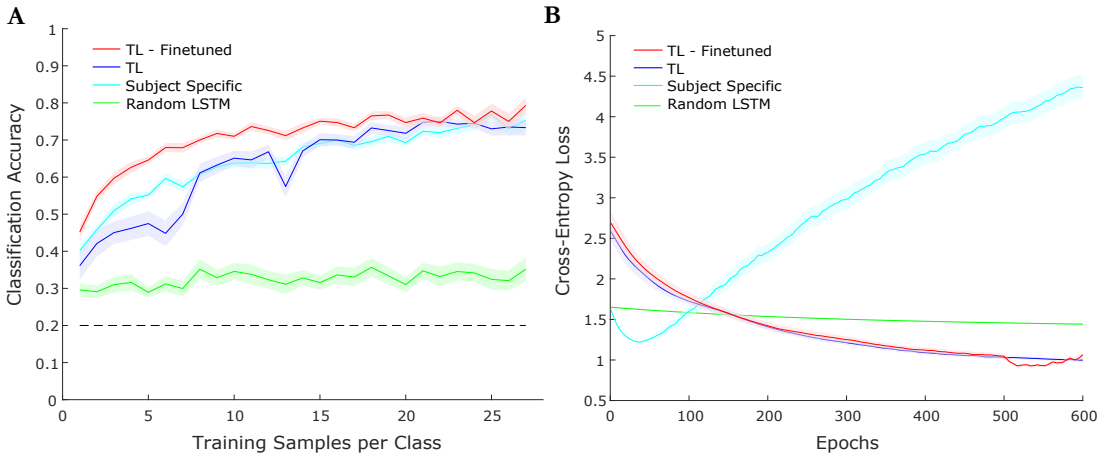

Figure 3.5: Transfer learning LSTM performance comparison. The TL-Finetuned, TL, and random LSTM all utilize Subject B as S_1 and Subject C as S_2. The subject specific model uses Subject C. (A) Accuracy as a function of the amount of training samples. (B) Cross-entropy loss across epochs, using 10 training samples per class. Error bars show standard error of mean averaging results from 20 random partitions of the shuffled data.

Detailing the performance for each TL and TL-Finetuned, we evaluate all 3 remaining subjects for each S_2. For the TL model, we found that the transfer between subjects is agnostic of S_1 specific training and performs similarly across all 3 subjects. The performance of TL-Finetuned is similarly evaluated, but has trends unique to TL. In particular, we observe that transferring from Subject A always provides the best results followed by transferring from Subject B. Accuracies using the maximum permissible training data for all four subjects comparing the two transfer learning approaches and the best linear model as well as the subject specific LSTM are reported in Table 3.2.

Visualizing Learned Representation

For the transfer learning models, we explored the stability of the affine layer and analyzed the learned LSTM weights between subjects. Examining the learned

Table 3.2: Summary comparing the average accuracy of transfer learning with subject specific training, using all training data across 20 random partitions of the shuffled data.

Model	Subject A	Subject B	Subject C	Subject D
Best Linear Classifier	0.53	0.61	0.79	0.65
Subject Specific	0.51	0.62	0.75	0.69
TL (S_1 = Subject A)	-	0.66	0.72	0.60
TL (S_1 = Subject B)	0.40	-	0.70	0.46
TL (S_1 = Subject C)	0.44	0.63	-	0.59
TL (S_1 = Subject D)	0.44	0.67	0.73	-
TL-Finetuned (S_1 = Subject A)	-	**0.71**	**0.82**	**0.70**
TL-Finetuned (S_1 = Subject B)	0.46	-	0.75	0.62
TL-Finetuned (S_1 = Subject C)	0.44	0.63	-	0.66
TL-Finetuned (S_1 = Subject D)	**0.53**	**0.71**	0.79	-

weights of the affine mapping layer, we can see that the mapping resembles a projection of electrodes to the appropriate brain regions. In Figure 3.6A, we show the learned affine transformations for two cases: a projection of sensorimotor-motor (R_{Motor}, $R_{Sensory}$) between the two subjects, and a projection also adding the occipital lobe electrodes ($R_{Occipital}$) for S_2. As the occipital lobe is involved with integrating visual information, it is not as informative for motor movement tasks. Therefore, we would expect, and in fact observe, there to be an affine transformation with weights for $R_{Occipital}$ closer to zero indicating an absence of electrodes in S_1 that contain similar information. Additionally, we can qualitatively observe the stability of the affine transform after the fixed $LSTM^{S_1}$ and $Softmax^{S_1}$ are relaxed. It is clear by looking between the left and right columns of Figure 3.6A that the learned weights from TL are a good representation and only require minor modification in TL-Finetuned.

Furthermore, exploring the use of LSTMs for transfer learning, a two-dimensional embedding of the LSTM output using t-SNE [41] on the training data was created. The t-SNE embedding is produced by first converting the pair-

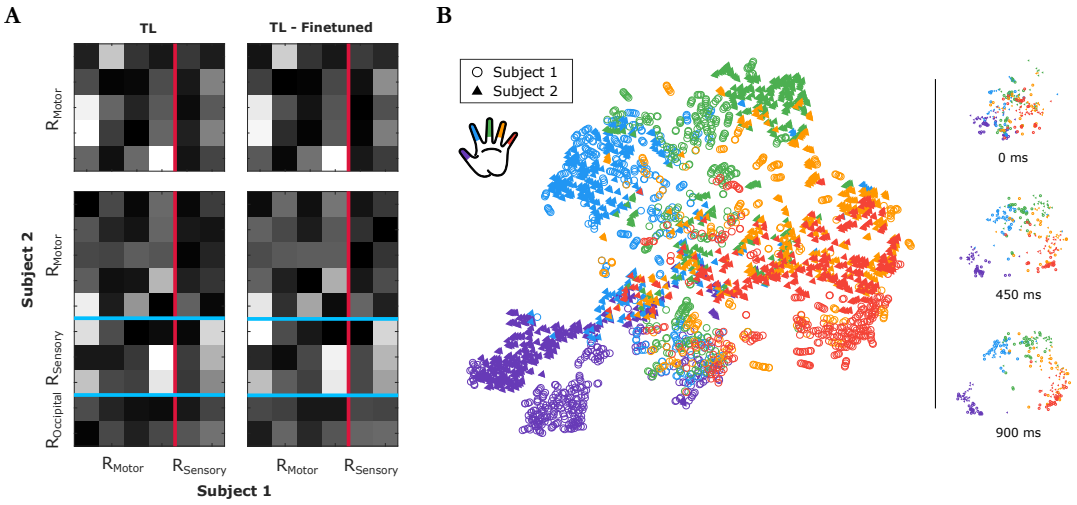

Figure 3.6: Visualizations of LSTM model parameters. (A) Weights of the learned affine mapping from Subject 2 (Subject C) to Subject 1 (Subject B). Electrodes mapped from motor to sensorimotor and sensorimotor + occipital to sensorimotor regions of S_2 to S_1, respectively. Both TL and TL-Finetuned weights are depicted. (B) Two-dimensional t-SNE embedding of the learned signal representation corresponding to different fingers for both subjects.

wise Euclidean distances, d_{ij}, in the high-dimensional space to joint probabilities $p_{ij} \propto \exp(-d_{ij}^2)$. We use the $LSTM^{S_1}$ outputs for S_1 data and $LSTM^{S_1}$ outputs for S_2 data after passing it through the learned affine layer on S_2. All the data between -300 ms to 900 ms, relative to behavioral onset, is embedded in the two-dimensional space. From the summary image, Figure 3.6B, it is clear that the model is able to separate classes for both subjects well and that the projections for both the subjects are clustered together. Identifying the source of the noise in the figure, we project the output at different time steps and see that the majority of the confusion occurs at the start of the sequences.

Acknowledgment

Sections in this chapter are adapted from work published and/or in preparation with Venkatesh Elango, Dr. Kai J Miller, and Dr. Vikash Gilja.

Chapter 4

Neural Interface Interaction

The development of long-term implantable systems and better decoders for neural prosthetic interfaces has the potential to allow for widespread adoption. As we develop these new interfaces, however, it is important to keep the ultimate goal of improving the quality of life for individuals who requires such systems in view. Reviewing the state-of-the-art neural interfaces, it is clear that the user interaction schemes are devised primarily for the maximization of performance [21, 7]. While important for establishing neural interfaces as a viable control interface, it does not take into account the cognitive burden or discomfort imposed on the user. In this chapter we define a set of usage modes and explore the effects different interaction schemes have on decoder performance while providing some insight into the advantages and limitations the decoder design decisions have.

4.1 Usage Modes

Examining the interactions we have daily with other people, technology, and our environment, there are a host of rich schemes that we utilize to fully experience the world [42]. Considering existing communication prostheses and other BCIs,

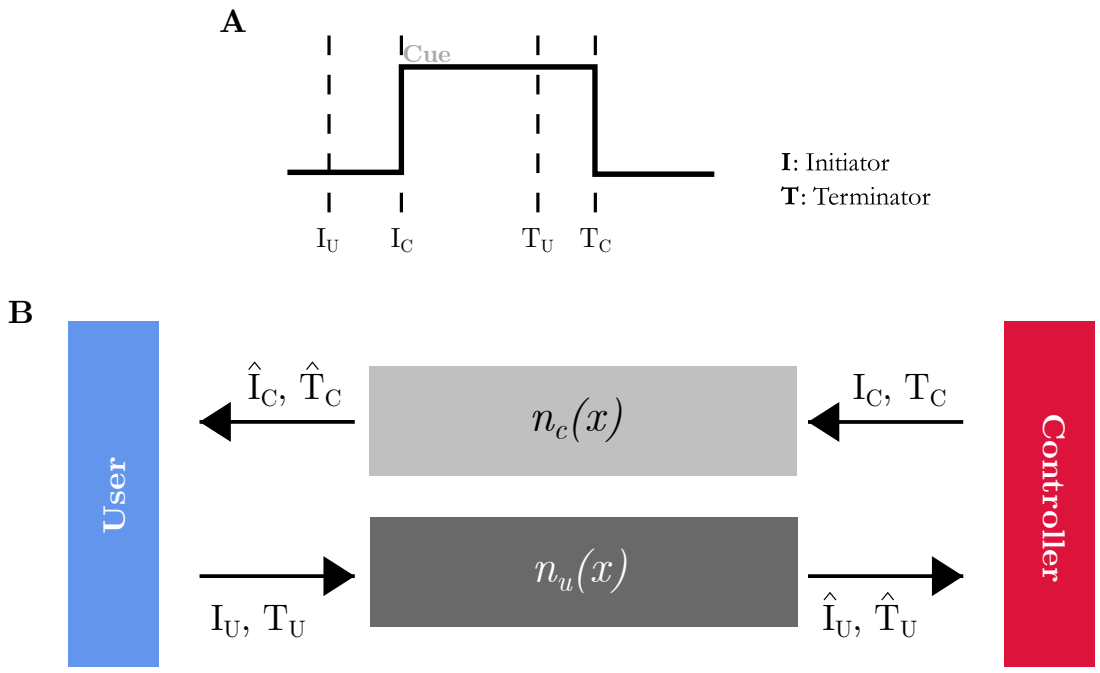

Figure 4.1: Timing breakdown of neural interface interactions.

we can simplify the interaction schemes between the user and a neural interface to being determined by two parameters - initiation (I) and termination (T). Each parameter has two possible values, user (U) or computer (C), that establish who is controlling the particular timing criteria. For example, a P300-like system that uses a rigid timing scheme would function on a $I_C T_C$, or equivalently CC and computer-computer, paradigm. On the opposite end of the spectrum, a $I_U T_U$, or equivalently UU and user-user, paradigm are representative of free-paced volitional control. Figure 4.1 depicts the timing characteristics with respect to a cue such as those present in the finger flexion dataset.

While $I_C T_C$ and $I_U T_U$ are the two most easily observed usage schemes, the other two hybrid modalities provide interesting design possibilities for exploration. The $I_C T_U$ control scheme, in particular, allows for a rigid initiation of behavior, but allows the user to control the duration of the intent signal. Though numerous

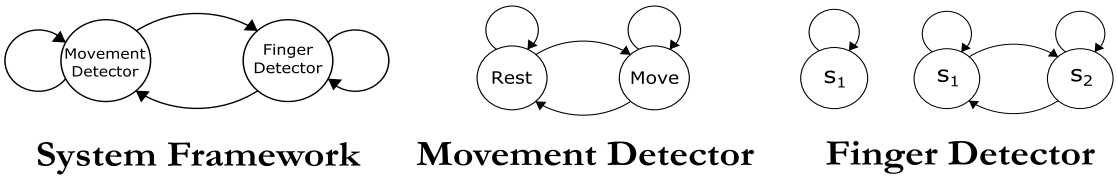

Figure 4.2: Hierarchical decoder system framework and associated movement detector and finger detector architectures using HMMs.

human-computer interaction (HCI) studies could be conducted exploring each of the proposed modes in detail, individuals could have different preferences for a range of applications. We present a structured approach to evaluating the model timing influence on performance which provides designers with useful insight when developing clinical systems.

4.2 Evaluation Methodology

Hidden Markov models were used to explore the interface timing characteristics. Different from the model proposed in Chapter 3, a hierarchical decoder is constructed to allow for explicit separation of the movement detector (Figure 4.2). While alternative HMM structures can provide similar single graph structures (i.e. hierarchical HMMs), the benefit of separating the functionality is to provide clear understanding of the timing characteristics while also mimicking gateway decoder structures. Alternative models were considered for use in evaluating interface interaction, however, the simple representation of sequence structures allows for ease of understanding with respect to physiological phenomenon compared to techniques like LSTMs.

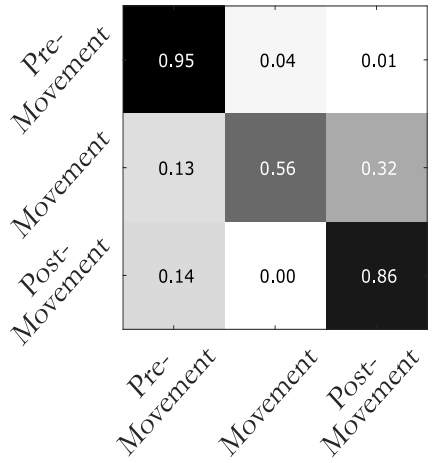

Figure 4.3: Segmented movement confusion matrix.

4.3 Results

The movement detector is the first component of the hierarchical decoder that is analyzed. Compared to linear decoders proposed in Chapter 2, the HMM model is separated into three states in an attempt to elucidate the performance characteristics between pre-movement and post-movement sequence information. Additionally, it is important to note the smaller integration window compared to that of the linear classifiers (150 ms versus 750 ms). Figure 4.3 shows that while pre-movement information is identifiable, the movement to post-movement transition is difficult to decode. Upon further examination of the ECoG signal over time, this is likely due to the decrease in power that occurs consistently across the duration of the trial as the behavior is repeated [18].

Examining the finger classification performance, there are three characteristics that we can explore: integration time, confidence, and class count. Figure 4.4 shows the integration time and confidence trade-offs. Looking first at the integration time, it is clear that the information rate is highest at the smallest integration

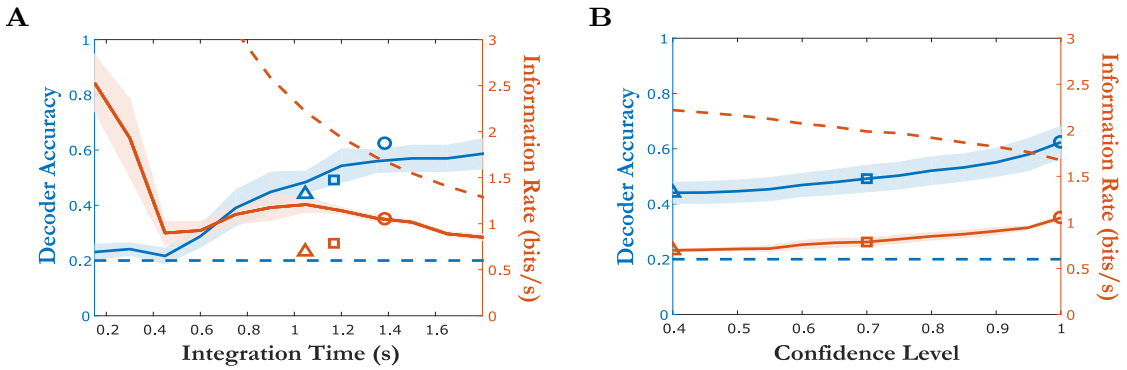

Figure 4.4: Evaluation of model timing characteristics. The dotted lines represent the theoretical max limits. (A) Model performance is evaluated with respect to the integration time used. (B) Model performance is evaluated with respect to decoder confidence level. The decoder accuracy and information rates are a function of all sequence lengths that achieve at least a fixed confidence. They are mapped as shapes onto integration time.

window at the cost of accuracy. This is opposite to the accuracy where the larger integration times achieve peak performance. Examining the confidence level of the decoder identifying a symbol, we can see both the decoder accuracy and information rate increase. Three specific points at 0.4, 0.7 and 1.0 confidences are selected and plotted as shapes on both Figure 4.4A, B. The range of integration time, interestingly, is 1 - 1.4 seconds indicating more time does not increase the decoder confidence, but rather be useful in correcting an incorrect classification.

Further exploring approaches to improving the decoder performance, the class and electrode region dependencies are explored (Figure 4.5). From Figure 4.5A it is clear that fewer symbols allow for matching theoretical capacity, where a greater number of symbols allows for more user commands at the cost of capacity. Examining the region dependence, a primary motivation for electrode selection in the models proposed in chapter Chapter 3 was the physiological significance. Using a 6-class single electrode LDA classifier for ranking, each electrode is evaluated with the HMM model. Looking at the performance across the different

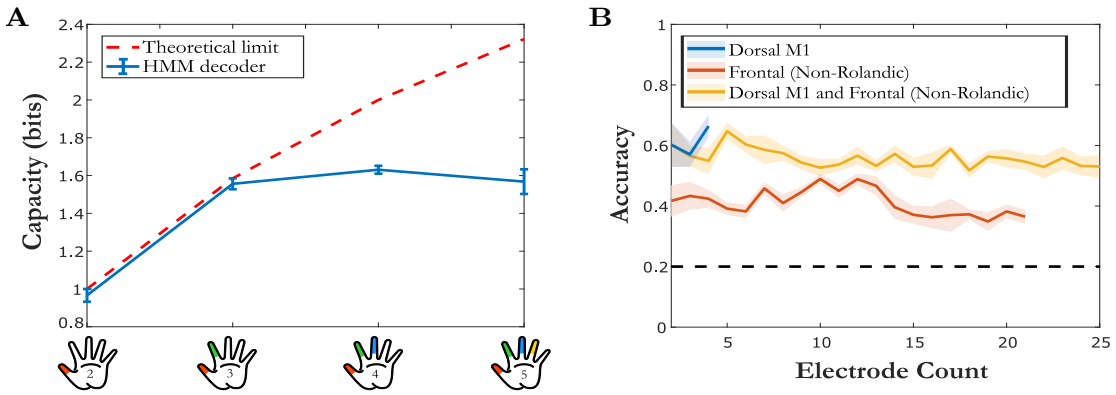

Figure 4.5: Decoder performance as a function of symbol count, electrode region, and electrode count. (A) Analysis of the symbol trade-offs. All unspecified fingers are grouped into a single class. (B) Region analysis of electrodes motivated by physiology.

regions, it is clear that the sensorimotor electrodes provide the best performance and with fewer sources. It is important to note that the ranking criteria does not take into account correlation between electrodes, but evaluates performance for each electrode independently. It is clear, however, that even when evaluating all implanted electrodes, it is important to identify the most informative electrodes.

4.4 Discussion

Reviewing the results from the linear classifiers using only a large onset window, we know it is possible to accurately decode the movement. Moving to smaller time bins and using sequence decoding techniques, the performance drops slightly, but is still able to classify movement well. This lends the use of a high-accuracy binary classifier for use in critical decision tasks or as a type of gateway decoder that selects modes. As a neural prosthetic designer, there are rich considerations that can be made from the results presented. Looking at information rate, if the primary goal is to maximize this metric, the smallest integration win-

dow yields the highest result. However, the resulting decoding accuracy is near chance due to a lack of informative signal being provided to the decoder. As one may expect, this approach would lead to a poor user experience. Depending on the location of the I_C and T_C in Figure 4.4, different rigid timing constraints can be considered. For example, the user could provide stimuli for as long as confidence is sub-threshold and be notified when the decoder is above the designated threshold. Evaluating this hybrid approach to timing where the interface and the user have an understanding, the confidence approach can provide a faster interface and achieves higher accuracy.

Examining the free-paced behavior scenarios, it is difficult to determine when a stimuli is no longer being provided. The primary issue is the early transition back to a rest state. With further improvement in the discriminability between these two states and the utilization of more powerful methods like LSTMs, it is conceivable that online performance can match that seen in offline evaluation. Furthermore, identifying the confusion between the classes, it is also possible to improve performance by either combining or eliminating the last two finger movements. And lastly, looking at the region filtered performances discussed in the Results, it is clear that it is not necessary to implant large grids to achieve good decoder performance; rather, using few electrodes placed over informative regions is preferred.

Acknowledgment

Sections in this chapter are adapted from work published and/or in preparation with Venkatesh Elango, Dr. Kai J Miller, and Dr. Vikash Gilja.

Chapter 5

Developing Neural Interfaces

The work presented thus far has been focused on improving the prosthetic interfaces with a particular focus on implantable devices that utilize ECoG. Unfortunately, the invasive nature of ECoG limits the experimental setups and collection duration [39, 40]. We have proposed techniques to overcome limitations specific to data availability and inter-patient variability in previous chapters, but the limitations imposed on the experimental setups have not been addressed. Fortunately, it is not necessary to implant subdural grids to measure neural responses. Non-invasive techniques can be utilized with varying levels of success. EEG is one such approach that is capable of measuring neural signals using electrodes that make contact with the scalp surface. Though signal quality is not comparable to ECoG, the setup allows for use outside of a clinical setting. Many interesting experiments have been conducted using EEG, but most must still be constrained to well controlled environments [43]. Recent work using immersive virtual reality and simulation chambers have allowed for more realistic experiments to be constructed, but have significant overhead involved with environmental setup and are therefore restricted to particular simulations such as driving or flight [44, 45]. In address-

ing this inability to easily construct experiments that are able to measure neural signals and other bio-sensory information in real-world settings, we developed a system that will bridge the gap between research and development of practical neural prosthetic interfaces.

5.1 Related Work

Electroencephalogram (EEG) systems have experienced a renewed interest for use in non-clinical studies. Though being deployed in large-scale studies, many of the advances have not been translated to substantial real-world applications. A major challenge is that the hardware and software used to make measurements. In particular, the low spatial resolution of EEG limits the amount of usable information that can be extracted from the dynamic recording environments. Additionally, the absence of methods to automatically extract user-environment interactions for tagging with EEG data introduces an immense overhead - having to manually tag events or limit experimental design by requiring the subjects to provide information during the experiments.

Most EEG research from the past few decades has been conducted under controlled laboratory environments as opposed to practical daily-use applications. Under these controlled conditions, EEG researchers have been able to control quadcopters [44], robots [46], and wheelchairs [47]. Unfortunately, the translation of this research to applications of EEG into the wild are not seen due to constraints imposed by the existing EEG decoders [17].

EEG research often studies event-related responses evoked or elicited by a visual or auditory stimulus. For real-world experiments with EEG, the stimulus onset is not measured or is ill-defined. A solution is to use saccadic eye movements

and fixations as the time-locking mechanism for analyzing naturalistic visual stimuli [48, 49, 50]. Hence, we must simultaneously record and synchronize EEG and eye-gaze data in real-world neuroimaging studies. For real-world experiments with EEG, there is also a need to pinpoint the stimulus that is causing the changes in EEG. Therefore, it is necessary to also record stimuli from the users' visual perspective for real-world experiments.

Exploring other sensing modalities that can influence neural decoding, we can look to emotion. For the analysis of emotional responses, recent research has shown that the heart plays a complementary role in the generation of emotions [51, 52]. This falsifies the wide ranging belief that the brain is solely responsible for the generation of emotion. Unfortunately, there is no existing system which can reliably sense and record EEG and electrocardiogram (ECG) synchronously in a mobile environment. Furthermore, ECG complicates the experimental design as subjects must wear an ECG belt or place several sensors on their chest. An alternative solution is to use photoplethsmogram (PPG) from commercially available devices that sense from the wrist. Such devices typically use low sampling rates to conserve power and therefore can only measure heart rate, as opposed to also measuring heart-rate variability (HRV).

5.2 System Overview

We propose a modular headset design to increase the flexibility and efficiency of multiple sensory data collection. The core control board was selected based on cost, community support, and compute capabilities. Arduinos and other ARM-based embedded systems were evaluated on their ability to achieve lower-bound frame-rates and collect data from multiple sensors in real-time using the

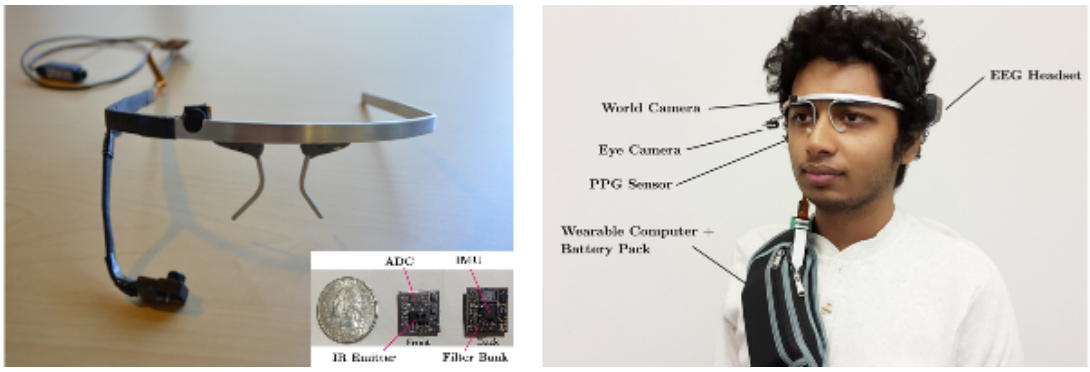

Figure 5.1: Proposed modular headset for capturing mult-modal sensory data. (Left) Headset for capturing eye-gaze and user view. (Left-bottom) PPG sensor. (Right) Complete system view of first design.

Lab-Streaming Layer (LSL) [53]. Ultimately, the Raspberry Pi 3 compute module (RPi3) was selected. The sensors that were selected for use in the first version are explored in detail in the following sections.

5.2.1 Electroencephalogram

Non-invasive EEG is used to collect neural signals from individuals. Any EEG system that is supported by LSL can be used in the proposed multi-modal bio-sensing framework. The Emotiv Epoc+ has a suitable trade-off between ease-of-use and performance. The Epoc+ allows for wireless collection of data that can be time-stamped and synchronized in real-time by the compute module. The sampling frequency of the system is on the lower end of new commercial systems at 128 Hz, but has 14 saline activated channels and a gyroscope allowing for collection of cleaner signals. Independent component analysis (ICA) [54, 55, 56] is used in real-time using the ORICA [57, 58] toolbox in Matlab to separate the sources of EEG recordings. For each of the independent components, the scalp map is plotted in real-time to better depict the source localization. ICA is also used to remove EEG artifacts due to eye blinks, muscle movement, and other sources.

Due to the uncomfortable nature of existing heart-rate and heart-rate variability sensors, a small PPG sensor (Figure 5.1) was developed that magnetically clips to the ear. The miniaturization was achieved by integrating a high-precision and high-sampling rate ADC to the sensor. Additionally, to eliminate noise, a third-order filter (band-pass 0.8 - 4 Hz) was also integrated such that only the filtered signals are digitized and transmitted to the control board. To also account for motion artifacts in the heart-rate signals, a 3-axis accelerometer was integrated into the board. The two data streams, once collected by the controller, are integrated using an adaptive noise cancellation algorithm (ANC) [59, 60]. Addressing the discomfort and bulk associated with existing systems, the device was developed to be mountable to the ear-lobe using magnets [59]. And because the system is low-profile and capable of resting behind the ear, more mobile studies can be conducted without the constrained natures of existing systems [61, 62].

5.2.2 Eye Gaze

The next sensor of the multi-modal system is a pair of cameras. One camera, an IR emitting device, is capable of accurately capturing the pupil location. A pupil-centering algorithm is integrated into the platform and is capable of maintaining the exact location even under perturbation. An algorithm developed by Pupil Labs for pupil detection and eye-gaze calibration is utilized [63]. The second integrated camera is a world-view camera. This wide-angle camera provides a view of what the wearer is seeing. With the information that is retrievable from both the pupil and the world cameras, it is possible to retrospectively reconstruct the full-view that the user was observing. The primary issue that stems from this type of video collection is the large amount of data that must be manually labeled. There are machine-learning tools that are capable of labeling video post-hoc [64, 65], but

limit the types of experiments that can be performed. To create a truly real-time system, the video can be streamed to a computer and processed using deep-learning libraries such as You Only Look Once (YOLO) that are capable of labeling 20 objects in real-time [66, 57]. By labeling exactly what the user is looking at and allowing labeled data to be accessible during the experiment, the rigidity of the experimental task can be relaxed allowing for more natural free-flowing behavior to be measured with minimally intrusive cues.

5.3 Evaluation

5.3.1 PPG

To quantify the performance of the miniaturized PPG sensor, different scenarios were considered to be representative of real-world scenarios. The baseline system for comparison is an EEG/ECG collection system from the Institute of Neural Engineering of Tsinghua University, Beijing, China. It is capable of measuring both signals at 1,000 Hz. Because the reference system takes measurements from electrodes placed near the heart, artifacts introduced from movements or other physiological responses are minimized. While taking measurements from the reference system, our PPG sensor is collecting the ECG signal from the ear lobe at a rate of 100 Hz. As both systems can be connected in parallel, they are synchronized using the lab-streaming layer and analyzed post-hoc.

The first experiment was a resting scenario - the user remained seated for a fixed period of two minutes. The PPG sensor data was compared to the reference with and without the adaptive noise cancellation filter. The second experiment was an active scenario where the user was instructed to walk in-place to simulate an active walking scenario. Here the data was also evaluated with and without the

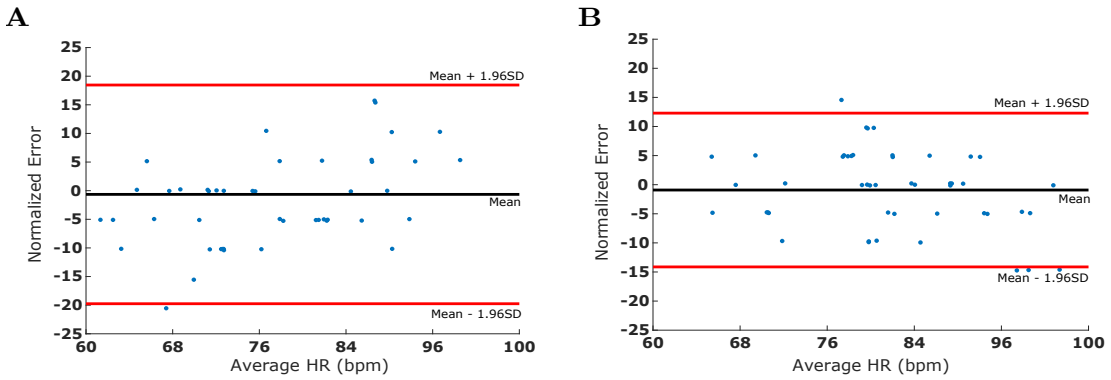

Figure 5.2: Characterization of PPG performance with respect to HR-strap during no movement. (A) PPG results without ANC and (B) PPG results with ANC.

adaptive noise cancellation filter.

A peak detection algorithm using minimum distance to next peak as one of the parameters for eliminating false peaks was used to calculate Heart Rate (HR) from ECG and PPG Data. 15-second trails were used to calculate the HR using the peak-detection algorithm. Figure 5.2 shows the normalized errors where a perfect HR estimation resulting in 0% error. Examining the results from the reference, ANC enabled, and ANC disabled signals, it is clear that the ANC enabled signals have the least noise and most closely match the reference signal. During resting measurement, the ANC disabled and ANC enabled signals were comparable (Figure 5.2). It is in active environments that having the ANC filtering provide a marked improvement in noise rejection (Figure 5.3).

5.3.2 Eye Gaze

The performance of the paired pupil and world view cameras was evaluated using a structured visual task to measure precision and accuracy during use. The user was seated 2-2.5 feet away from a computer monitor such that $>$ 90% of

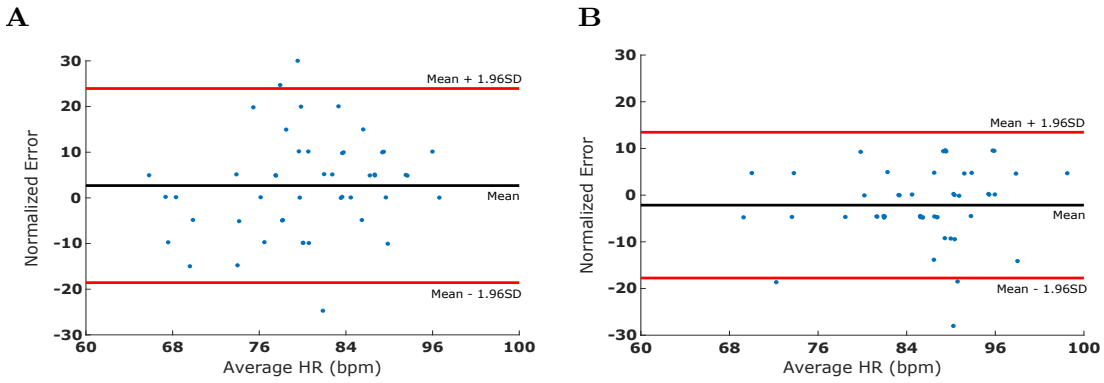

Figure 5.3: Characterization of PPG performance with respect to HR-strap during movement. (A) PPG results without ANC and (B) PPG results with ANC.

the world camera view was composed of the task screen. Both cameras were streamed at 30 fps. The participant was instructed to fix their head movement and only move their eyes to gaze at static targets that appeared on the screen for the first task. A calibration step where 9 targets appeared in a regular pattern on the screen calibrated the gaze marker. Immediately following the calibration process, a series of 20 unique targets spread uniformly across the visual space are randomly cycled. This process was followed by 30 seconds of rest where the user was asked to move their head around without removing the headset. This action was designed to simulate the active movement scenarios while wearing the headset. Next, the participant is instructed to return to a preferred position and maintain head position. Twenty new uniformly distributed points are cycled on screen to measure the precision and accuracy of the eye-tracking system after active use. This process was repeated for a total of three trials per subject.

Examining the results, we observe that the accuracy and precision of the eye gaze setup does not drift significantly from the expected output. The accuracy is measured as the average angular offset (distance in degrees of the visual angle)

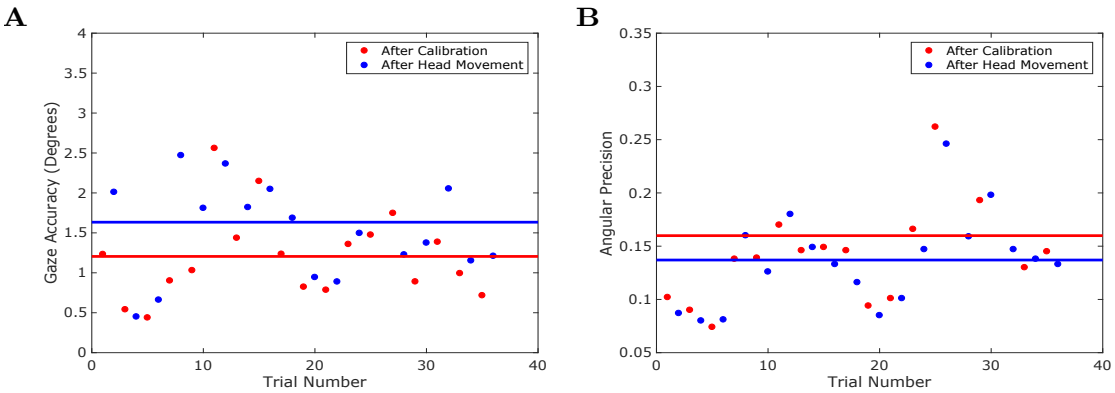

Figure 5.4: Eye gaze (A) accuracy and (B) precision comparisons before and after movement.

between fixation locations and the corresponding fixation targets (Figure 5.4A). The precision is measured as the root-mean-square of the angular distance (degree of visual angle) between successive samples during a fixation (Figure 5.4B). Compared to literature, the gaze accuracy drift of 0.42 degrees is significantly less than the 1-2 degree drift found in commercial systems [67, 68]. The precision, on the other hand, experiences only a 0.2 degree shift post movement, indicating a minimal angular distance shift.

5.4 Discussion

The results presented above are focused on assessing the capabilities of the multi-modal headset. Ongoing work has been focused on developing custom EEG electrodes, improving packaging efficiency of the electronics, and securing all components onto a single headset. Having established eye gaze and PPG as stable sensors, studies can be conducted using existing mobile EEG headsets like the Emotiv Epoc+.

Acknowledgment

Sections in this chapter are adapted from work published and/or in preparation with Siddharth, Dr. Tzyy-Ping Jung, and Dr. Terrence J. Sejnowski.

Chapter 6

Conclusion

Three major themes are presented in this thesis with a few key takeaways. The major themes addressing current problems with developing practical neural prosthetic interfaces are: improving decoder performance, developing for the user through mindful interfaces, and creating tools to enable researchers to conduct research that more easily translates to clinical or commercial applications.

Reviewing the first theme, little existing work has focused on efficient modeling of time-varying dynamics of neural signals. In this thesis, we have shown that LSTMs can model the variation within a neural sequence and are a good alternative to state-of-the-art linear decoders. Even with low sample counts and comparably greater number of parameters, the model is able to extract useful information without overfitting. Moreover, LSTMs provide a robust framework that is capable of scaling with large sample counts as opposed to the the limited scalability provided by linear classifiers. Considering the limitations imposed on our model by stopping at a fixed evaluation epoch, it would be possible to further boost performance by utilizing early stopping with a validation set. And while the input features were selected from empirical observations made in previous studies, the

results could be improved by extracting the features in an unsupervised manner using autoencoders [69] or by training the decoder end-to-end using convolutional LSTMs [70, 71].

Furthermore, having established the LSTM as a good approach for neural decoding, we explored utilizing the model in transfer learning scenarios. Exploring a less constrained model where the LSTM weights are relaxed, the performance is shown to exceed that of both the subject specific training and the best linear decoder models in both the average scenario for all mappings as well as the best patient mapping. This robustness against subject specific neural dynamics even when only the affine transform is learned indicates that the LSTM is capable of extracting useful information that generalizes to the new patient with limited impact due to the original subject's relative performance. When exploring the trade-offs between TL and TL-Finetuned, the latter provides performance that exceeds the current state-of-the-art models with fewer subject-specific training samples where the former is able to achieve comparable performance. However, because TL only requires training of the affine layer, the computation is less expensive than the TL-Finetuned. From Figure 3.5, it could be seen when learning only the affine transformation the cross-entropy loss is still decreasing after 500 epochs suggesting that with better optimization methods, the TL model by itself may outperform the subject-specific model. This motivates the statement that the LSTM is capable of extracting a representation of the neural variability between behaviors that generalizes across patients. While this may be specific to the behavior being measured, it posits potential scenarios for using sequence transfer learning.

The next theme explores the user interactions that should be accounted for when developing neural prostheses. A hierarchical decoder construction using HMMs was used to explore these timing constraints using a stereotypical finger

flexion experiment. While proposing a framework to categorize the interaction schemes between the user and computer or controller, different trade-offs are presented. The primary consideration is the rigid timing constrain posed by standard BCIs - by using a flexible timing scheme, a BCI could be constructed that has less accuracy (a function of the decoder), but provides greater utility and comfort to the user. We also propose a hybrid approach that utilizes the decoder confidence in unison with the user's input to control the output. While these are two analyses that are performed to explore the timing and interface control trade-offs, other important considerations like the symbol count and number of electrodes are also explored and discussed. Ultimately, it is important to keep the user in mind when developing neural prosthetic interfaces - the goal is to improve their quality of life through better decoders and more usable interfaces.

The primary focus of the final theme was to develop a multi-modal platform that is capable of integrating multiple sensors into a comfortable and wireless system that enables researchers to conduct experiments that are more realistic and translate more readily to clinical and commercial applications. The core system that has been tested and validated against established measurement systems are the eye gaze tracking, heart rate, and heart rate variability sensors. Paired with commercial EEG solutions like the Emotiv Epoc+, the proposed solution provides unmatched mobility and flexibility to explore interesting neural phenomenon in unstructured environments. The perspective gained from these analyses provide better insights to practical neural interface design.

Appendix A

Hidden Markov Models

A.1 Index Notation

Initialization

$$O^1, O^2, ..., O^K - K \text{ sequences}$$

$$o_1^k, o_2^k, ..., o_T^k - \text{Sequence } k$$

Initial State Probabilities

$$N - \text{Number of states}$$

$$\pi_i - \text{Initial state probabilities}$$

$$b_i(O) = P(O; \mu_i, \Sigma_i) - \text{Gaussian Emission}$$

$$(\mu_i, \Sigma_i) - \text{Set of } N \text{ mean, covariance } \forall i \in \{1, 2, ..., N\}$$

$$[A]_{ij} - \text{Transition probabilities } \forall i \ \forall j \in \{1, 2, ..., N\}$$

$$S_t - \text{State at time } t, \ \forall t \in \{1, 2, ..., T-1\}$$

$$\pi_i = \begin{cases} 1 & i = 1 \\ 0 & i \neq 1 \end{cases}$$

Define:

$$\alpha_{it}^k \triangleq P(o_1^k, ..., o_t^k, S_t = i)$$
$$\beta_{it}^k \triangleq P(o_{t+1}^k, ..., o_T^k | S_t = i)$$
$$l_{it}^k \triangleq \max_{S_1,...,S_{t-1}} \log P(S_1, S_2, ..., S_{t-1}, S_t = i, o_1, ..., o_t)$$

The forward-backward algorithm can be stated as:

$$\begin{aligned}
\alpha_{j1}^k &= \pi_j b_j(o_1^k) && \forall j \in \{1, 2, ..., N\} \\
\alpha_{jt+1} &= \sum_{i=1}^{N} \alpha_{it}^k A_{ij} b_j(o_{t+1}^k) && \forall t \in \{1, 2, ..., T-1\} \\
P(O^k; \lambda) &= \sum_{i=1}^{N} \alpha_{iT}^k \\
\beta_{iT}^k &= 1 && \forall i \in \{1, 2, ..., N\} \\
\beta_{it}^k &= \sum_{j=1}^{N} b_j(o_{t+1}^k) \beta_{jt+1}^k A_{ij} && \forall t \in \{T-1, ..., 2, 1\}
\end{aligned}$$

Followed by Expectation-Maximization.

E-step:

$$P(S_t^k = i, S_{t+1}^k = j | o_1^k, ..., o_T^k) = \frac{\alpha_{it}^k A_{ij} b_j(o_{t+1}^k) \beta_{jt+1}^k}{\sum_{i=1}^{N} \alpha_{iT}^k}$$

$$P(S_t^k = i | o_1^k, ..., o_T^k) = \sum_{j=1}^{N} P(S_t^k = i, S_{t+1}^k = j | o_1^k, ..., o_T^k)$$

M-step:

$$A_{ij} \leftarrow \frac{\sum_{k=1}^{K} \sum_{t=1}^{T-1} P(S_t^k = i, S_{t+1}^k = j | o_1^k, ..., o_T^k)}{\sum_{k=1}^{K} \sum_{t=1}^{T-1} P(S_t^k = i | o_1^k, ..., o_T^k)}$$

$$\mu_i \leftarrow \frac{\sum_{k=1}^{K} \sum_{t=1}^{T} P(S_t^k = i | o_1^k, ..., o_T^k) o_t^k}{\sum_{k=1}^{K} \sum_{t=1}^{T} P(S_t^k = i | o_1^k, ..., o_T^k)}$$

$$\Sigma_i \leftarrow \frac{\sum_{k=1}^{K} \sum_{t=1}^{T} P(S_t^k = i | o_1^k, ..., o_T^k)(o_t^k - \mu_i)(o_t^k - \mu_i)^T}{\sum_{k=1}^{K} \sum_{t=1}^{T} P(S_t^k = i | o_1^k, ..., o_T^k)}$$

Prediction on new data can be performed using the Viterbi decode algorithm.

$$l_{i1}^k = \log \pi_i + \log b_i(o_1^k) \qquad \forall i \in \{1, 2, ..., N\}$$

$$l_{jt+1}^k = \log b_j(o_{t+1}^k) + \max_i (l_{it}^k + log A_{ij}) \quad \forall j \in \{1, 2, ..., N\}, \forall t \in \{1, 2, ..., T-1\}$$

State Transitions:

$$\Phi^k_{t+1}(j) = \arg\max_i (l^k_{iT} + \log A_{ij}) \qquad \forall j \in \{1, 2, ..., N\},\ \forall t \in \{1, 2, ..., T-1\}$$

$$S^k_T = \arg\max_i (l^k_{iT})$$

$$S^k_t = \Phi^k_{t+1}(S^k_{t+1}) \qquad \forall t \in \{T-1, ..., 1\}$$

A.2 Vector Notation

The subscript comma and $\times$ are used interchangeably to state dimensionality, the following dimensions are:

$$N - States$$

$$K - Sequences$$

$$T - Time$$

$$F - Features$$

And : indicates a slice across all values in the specified dimension. The following variables denote:

$$O_{K,T,F} - \text{Sequences}$$

$$A_{N\times N} - \text{Transition probabilities}$$

$$\pi_{N,1} - ([1,0,...,0]^T)_{N\times 1}\text{; Initial state probabilities}$$

$$(\mu_F, \Sigma_{F\times F})_N - \text{Set of means and covariances}$$

$$B_{K,N,T} - B_{k,n,t} = P(O_{k\times t}, \mu_n, \Sigma_n)$$

$$\alpha_{K,N,T} - \text{Forward}$$

$$\beta_{K,N,T} - \text{Backward}$$

$$P(O)_{K\times 1} - \text{Sequence Probability}$$

The forward-backward algorithm can be stated as:

$$\alpha_{:,:,1} = \pi \odot B_{:,1,:}$$

$$\alpha_{:,:,t+1} = \alpha_{:,:,t} \times A \odot B_{:,:,t+1} \qquad \forall t \in \{1,2,...,T-1\}$$

$$P(O) = \sum_{i=0}^{T} \alpha_{:,:,i}$$

$$\beta_{:,:,T} = \mathbf{1}_{K\times N}$$

$$\beta_{:,:,t} = A \times (B_{:,:,t+1} \odot \beta_{:,:,t+1})$$

Followed by Expectation-Maximization.

E-step:

$$
\begin{aligned}
P(S_t, S_{t+1})_K - &\text{Tensor(K, T-1, N, N)} \\
P(S_t) - &\text{Tensor(K, T-1, N)} \\
P(S_t, S_{t+1}) &= \frac{\alpha_{:,:,t} \times A \odot B_{:,:,t+1} \odot \beta_{:,:,t+1}}{P(0)} \qquad \forall t \in \{1, 2, ..., T-1\} \\
P(S_t) &= \sum_N P(S_t, S_{t+1}))
\end{aligned}
$$

M-step:

$$
\begin{aligned}
A &= \frac{\sum_K \sum_T P(S_t, S_{t+1})}{\sum_K \sum_T P(S_t)} \\
\mu_i &= \frac{\sum_K \sum_T P(S_t)_{:,:,i} \odot O}{\sum_K \sum_T P(S_t)} \qquad \forall i \in \{1, 2, ..., N\} \\
\Sigma_i &= \frac{\sum_K \sum_T P(S_t)_{:,:,i} \odot (O - \mu_i) \times (O - \mu_i)^T}{\sum_K \sum_T P(S_t)} \qquad \forall i \in \{1, 2, ..., N\}
\end{aligned}
$$

Inference is identical to index notation as each sequence is evaluated independently.

A.3 Psuedocode

Algorithm 1 Hidden Markov Model - Gaussian Emission

procedure LEARNING($seqs$, π)
  $y \leftarrow N(\mu, \Sigma).\text{PDF}(seqs)$
Forward:
  $\alpha_0 = \pi \times y_0$
  for $i := 1$ **to** t **step** 1 **do**
    $\alpha_i \leftarrow \alpha_{i-1} \times P \odot y_i$
Backward:
  $\beta_t \leftarrow \mathbf{1}$
  for $i := 0$ **to** $t-1$ **step** 1 **do**
    for $s := 0$ **to** S **step** 1 **do**
      $P'_s \leftarrow P \times \text{diag}(y_{t-i,s})$
    $\beta_{t-i-1} \leftarrow P' \times \beta_{t-i}$
Expectation:
  $\gamma = (\alpha \odot \beta)/\text{sum}(\alpha_T)$
  $\xi = (P \odot \alpha_{0\ldots t-1} \times (y \times \beta)_{1\ldots t}) \,/\text{sum}(\alpha_{t,N})$
Maximization:
  $P \leftarrow \text{sum}(\xi_{T,S})/\text{sum}(\gamma_{0\ldots t-1,S})$
  $\mu' \leftarrow \gamma \times seqs$
  $\Sigma' \leftarrow \text{tile}(\gamma, F) \odot (\text{tile}(seqs, N) - \text{tile}(\mu, T)) \times \text{tile}(\gamma, F)$
  $\mu \leftarrow \text{sum}(\mu'_S)/\text{sum}(\gamma_{T,S})$
  $\Sigma \leftarrow \text{sum}(\Sigma'_S)/\text{sum}(\gamma_{T,S})$
end procedure

procedure INFERENCE($seqs$)
  $y \leftarrow N(\mu, \Sigma).\text{PDF}(seqs)$
  $\alpha_0 = \pi \times y_0$
  for $i := 1$ **to** t **step** 1 **do**
    $\alpha_i \leftarrow \alpha_{i-1} \times P \odot y_i$
  $l = \text{sum}(\alpha_{t,N})$
  $scores'_0 \leftarrow \text{tile}(\log y_0 + \log \pi, N)$
  for $i := 1$ **to** t **step** 1 **do**
    $scores'_i \leftarrow \text{tile}((\max((ss_{i-1} + \log P)_N) + \log y_i), N)$
  $states' \leftarrow \text{argmax}(scores'_N)$
  $scores \leftarrow \max(scores'_N)$
  $states_t \leftarrow \text{argmax}(scores_{t,N})$
  for $i := 0$ **to** $t-1$ **step** 1 **do**
    $states_{t-i-1} \leftarrow \text{gather}(states'_{t-i}, states_{t-i})$
end procedure

Appendix B

Long Short Term-Memory Networks

The defining equations for the LSTM cell are defined as

$$i_t = \text{sigmoid}(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \tag{B.1}$$

$$j_t = \tanh(W_{xj}x_t + W_{hj}h_{t-1} + b_j) \tag{B.2}$$

$$f_t = \text{sigmoid}(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \tag{B.3}$$

$$o_t = \text{sigmoid}(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \tag{B.4}$$

$$c_t = c_{t-1} \odot f_t + i_t \odot j_t \tag{B.5}$$

$$h_t = \tanh(c_t) \odot o_t \tag{B.6}$$

where the W terms are weight matrices, the b terms are biases, and $\odot$ represents Hadamard multiplication per the [29] paper. To enable gradient flow, the forget gate bias term is initialized to 1 [72, 73].

Bibliography

[1] Peter Howitt, Ara Darzi, Guang-Zhong Yang, Hutan Ashrafian, Rifat Atun, James Barlow, Alex Blakemore, Anthony MJ Bull, Josip Car, Lesong Conteh, Graham S Cooke, Nathan Ford, Simon AJ Gregson, Karen Kerr, Dominic King, Myutan Kulendran, Robert A Malkin, Azeem Majeed, Stephen Matlin, Robert Merrifield, Hugh A Penfold, Steven D Reid, Peter C Smith, Molly M Stevens, Michael R Templeton, Charles Vincent, and Elizabeth Wilson. Technologies for global health. *The Lancet*, 380(9840):507–535, 2012.

[2] H Goujon, X Bonnet, P Sautreuil, M Maurisset, L Darmon, P Fode, and F Lavaste. A functional evaluation of prosthetic foot kinematics during lower-limb amputee gait. *Prosthetics and orthotics international*, 30(2):213–223, 2006.

[3] Leigh R Hochberg, Daniel Bacher, Beata Jarosiewicz, Nicolas Y. Masse, John D Simeral, Joern Vogel, Sami Haddadin, Jie Liu, Sydney S Cash, Patrick van der Smagt, and John P Donoghue. Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature*, 485(7398):372–375, 2012.

[4] He Huang, Fan Zhang, Levi J Hargrove, Zhi Dou, Daniel R Rogers, and Kevin B Englehart. Continuous locomotion-mode identification for prosthetic legs based on neuromuscular–mechanical fusion. *IEEE Transactions on Biomedical Engineering*, 58(10):2867–2875, 2011.

[5] Christian Cipriani, Franco Zaccone, Silvestro Micera, and M Chiara Carrozza. On the shared control of an emg-controlled prosthetic hand: analysis of user–prosthesis interaction. *IEEE Transactions on Robotics*, 24(1):170–184, 2008.

[6] Francis HY Chan, Yong-Sheng Yang, FK Lam, Yuan-Ting Zhang, and Philip A Parker. Fuzzy emg classification for prosthesis control. *IEEE transactions on rehabilitation engineering*, 8(3):305–311, 2000.

[7] Y. Wang, Y. t. Wang, and T. p. Jung. Visual stimulus design for high-rate ssvep bci. *Electronics letters*, 46(15):1057–1058, 2010.

[8] Laura Acqualagna, Matthias Sebastian Treder, Martijn Schreuder, and Benjamin Blankertz. A novel brain-computer interface based on the rapid serial visual presentation paradigm. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pages 2686–2689. IEEE, 2010.

[9] Dean J Krusienski, Eric W Sellers, François Cabestaing, Sabri Bayoudh, Dennis J McFarland, Theresa M Vaughan, and Jonathan R Wolpaw. A comparison of classification techniques for the p300 speller. *Journal of neural engineering*, 3(4):299, 2006.

[10] Cynthia A Chestek, Vikash Gilja, Christine H Blabe, Brett L Foster, Krishna V Shenoy, Josef Parvizi, and Jaimie M Henderson. Hand posture classification using electrocorticography signals in the gamma band over human sensorimotor brain areas. *Journal of neural engineering*, 10(2):026002, 2013.

[11] Jennifer L Collinger, Brian Wodlinger, John E Downey, Wei Wang, Elizabeth C Tyler-Kabara, Douglas J Weber, Angus JC McMorland, Meel Velliste, Michael L Boninger, and Andrew B Schwartz. High-performance neuroprosthetic control by an individual with tetraplegia. *The Lancet*, 381(9866):557–564, 2013.

[12] Takufumi Yanagisawa, Masayuki Hirata, Youichi Saitoh, Haruhiko Kishima, Kojiro Matsushita, Tetsu Goto, Ryohei Fukuma, Hiroshi Yokoi, Yukiyasu Kamitani, and Toshiki Yoshimine. Electrocorticographic control of a prosthetic arm in paralyzed patients. *Annals of neurology*, 71(3):353–361, 2012.

[13] Vikash Gilja, Paul Nuyujukian, Cindy A Chestek, John P Cunningham, M Yu Byron, Joline M Fan, Mark M Churchland, Matthew T Kaufman, Jonathan C Kao, Stephen I Ryu, and Krishna V Shenoy. A high-performance neural prosthesis enabled by control algorithm design. *Nature neuroscience*, 15(12):1752–1757, 2012.

[14] John D Yeager, Derrick J Phillips, David M Rector, and David F Bahr. Characterization of flexible ecog electrode arrays for chronic recording in awake rats. *Journal of neuroscience methods*, 173(2):279–285, 2008.

[15] Alexander M Aravanis, Li-Ping Wang, Feng Zhang, Leslie A Meltzer, Murtaza Z Mogri, M Bret Schneider, and Karl Deisseroth. An optical neural interface: in vivo control of rodent motor cortex with integrated fiberoptic and optogenetic technology. *Journal of neural engineering*, 4(3):S143, 2007.

[16] Mariska J Vansteensel, Elmar GM Pels, Martin G Bleichner, Mariana P Branco, Timothy Denison, Zachary V Freudenburg, Peter Gosselaar, Sacha Leinders, Thomas H Ottens, Max A Van Den Boom, Peter C Van Rijen,

Erik J Aarnoutse, and Nick F Ramsey. Fully implanted brain–computer interface in a locked-in patient with als. *New England Journal of Medicine*, 375(21):2060–2066, 2016.

[17] Theresa M Vaughan, Jonathan R Wolpaw, and Emanuel Donchin. Eeg-based communication: Prospects and problems. *IEEE transactions on rehabilitation engineering*, 4(4):425–430, 1996.

[18] Kai J Miller, Dora Hermes, Christopher J Honey, Adam O Hebb, Nick F Ramsey, Robert T Knight, Jeffrey G Ojemann, and Eberhard E Fetz. Human motor cortical activity is selectively phase-entrained on underlying rhythms. *PLoS Comput Biol*, 8(9):e1002655, 2012.

[19] Kai J Miller and Jeffrey G Ojemann. A library of human electrocorticographic data and analyses. In *2016 Neuroscience Meeting Planner*. Society for Neuroscience, 2016.

[20] Kai J Miller, Eric C Leuthardt, Gerwin Schalk, Rajesh PN Rao, Nicholas R Anderson, Daniel W Moran, John W Miller, and Jeffrey G Ojemann. Spectral changes in cortical surface potentials during motor movement. *Journal of Neuroscience*, 27(9):2424–2432, 2007.

[21] Guy Hotson, David P McMullen, Matthew S Fifer, Matthew S Johannes, Kapil D Katyal, Matthew P Para, Robert Armiger, William S Anderson, Nitish V Thakor, Brock A Wester, and Nathan E Crone. Individual finger control of a modular prosthetic limb using high-density electrocorticography in a human subject. *Journal of neural engineering*, 13(2):026017, 2016.

[22] Zuoguan Wang, Gerwin Schalk, and Qiang Ji. Anatomically constrained decoding of finger flexion from electrocorticographic signals. In *Advances in neural information processing systems*, pages 2070–2078, 2011.

[23] Tobias Wissel, Tim Pfeiffer, Robert Frysch, Robert T Knight, Edward F Chang, Hermann Hinrichs, Jochem W Rieger, and Georg Rose. Hidden markov model and support vector machine based decoding of finger movements using electrocorticography. *Journal of neural engineering*, 10(5):056020, 2013.

[24] Lawrence R Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[25] Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. Speech recognition with deep recurrent neural networks. In *Acoustics, speech and signal processing (icassp), 2013 ieee international conference on*, pages 6645–6649. IEEE, 2013.

[26] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112, 2014.

[27] Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*, 2014.

[28] Pouya Bashivan, Irina Rish, Mohammed Yeasin, and Noel Codella. Learning representations from eeg with deep recurrent-convolutional neural networks. *arXiv preprint arXiv:1511.06448*, 2015.

[29] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[30] Joe Yue-Hei Ng, Matthew Hausknecht, Sudheendra Vijayanarasimhan, Oriol Vinyals, Rajat Monga, and George Toderici. Beyond short snippets: Deep networks for video classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4694–4702, 2015.

[31] Andrew M Dai and Quoc V Le. Semi-supervised sequence learning. In *Advances in Neural Information Processing Systems*, pages 3079–3087, 2015.

[32] Nitish Srivastava, Elman Mansimov, and Ruslan Salakhudinov. Unsupervised learning of video representations using lstms. In *International Conference on Machine Learning*, pages 843–852, 2015.

[33] Yarin Gal and Zoubin Ghahramani. A theoretically grounded application of dropout in recurrent neural networks. In *Advances in Neural Information Processing Systems*, pages 1019–1027, 2016.

[34] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, pages 41–48. ACM, 2009.

[35] Wojciech Zaremba and Ilya Sutskever. Learning to execute. *arXiv preprint arXiv:1410.4615*, 2014.

[36] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[37] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? In *Advances in neural information processing systems*, pages 3320–3328, 2014.

[38] Venkatesh Elango. *Sequence Learning for Brain Computer Interfaces.* M.S. Thesis, University of California, San Diego, 2017.

[39] JD Simeral, Sung-Phil Kim, MJ Black, JP Donoghue, and LR Hochberg. Neural control of cursor trajectory and click by a human with tetraplegia 1000 days after implant of an intracortical microelectrode array. *Journal of neural engineering*, 8(2):025027, 2011.

[40] Chethan Pandarinath, Paul Nuyujukian, Christine H Blabe, Brittany L Sorice, Jad Saab, Francis R Willett, Leigh R Hochberg, Krishna V Shenoy, and Jaimie M Henderson. High performance communication by people with paralysis using an intracortical brain-computer interface. *eLife*, 6:e18554, 2017.

[41] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(Nov):2579–2605, 2008.

[42] SG Mason, A Bashashati, M Fatourechi, KF Navarro, and GE Birch. A comprehensive survey of brain interface technology designs. *Annals of biomedical engineering*, 35(2):137–169, 2007.

[43] Peter Aspinall, Panagiotis Mavros, Richard Coyne, and Jenny Roe. The urban brain: analysing outdoor physical activity with mobile eeg. *British journal of sports medicine*, pages bjsports–2012, 2013.

[44] Karl LaFleur, Kaitlin Cassady, Alexander Doud, Kaleb Shades, Eitan Rogin, and Bin He. Quadcopter control in three-dimensional space using a noninvasive motor imagery-based brain–computer interface. *Journal of neural engineering*, 10(4):046003, 2013.

[45] QiBin Zhao, LiQing Zhang, and Andrzej Cichocki. Eeg-based asynchronous bci control of a car in 3d virtual reality environments. *Chinese Science Bulletin*, 54(1):78–87, 2009.

[46] Christian J Bell, Pradeep Shenoy, Rawichote Chalodhorn, and Rajesh PN Rao. Control of a humanoid robot by a noninvasive brain–computer interface in humans. *Journal of neural engineering*, 5(2):214, 2008.

[47] Tom Carlson and Jose del R Millan. Brain-controlled wheelchairs: a robotic architecture. *IEEE Robotics & Automation Magazine*, 20(1):65–73, 2013.

[48] Hélène Devillez, Nathalie Guyader, and Anne Guérin-Dugué. An eye fixation–related potentials analysis of the p300 potential for fixations onto a target object when exploring natural scenes. *Journal of vision*, 15(13):20–20, 2015.

[49] Juan E Kamienkowski, Matias J Ison, Rodrigo Quian Quiroga, and Mariano Sigman. Fixation-related potentials in visual search: A combined eeg and eye tracking studyfixation-related potentials in visual search. *Journal of vision*, 12(7):4–4, 2012.

[50] Laura Acqualagna and Benjamin Blankertz. Gaze-independent bci-spelling using rapid serial visual presentation (rsvp). *Clinical Neurophysiology*, 124(5):901–908, 2013.

[51] EE Van der Wall and WH Van Gilst. Neurocardiology: close interaction between heart and brain. *Neth Heart J*, 21(2):51–52, 2013.

[52] Martin A Samuels. The brain–heart connection. *Circulation*, 116(1):77–84, 2007.

[53] Kothe Christian. Lab streaming layer (lsl), 2015.

[54] Scott Makeig, Anthony J Bell, Tzyy-Ping Jung, and Terrence J Sejnowski. Independent component analysis of electroencephalographic data. *Advances in neural information processing systems*, pages 145–151, 1996.

[55] Scott Makeig, Tzyy-Ping Jung, Anthony J Bell, Dara Ghahremani, and Terrence J Sejnowski. Blind separation of auditory event-related brain responses into independent components. *Proceedings of the National Academy of Sciences*, 94(20):10979–10984, 1997.

[56] Anthony J Bell and Terrence J Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural computation*, 7(6):1129–1159, 1995.

[57] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338, 2010.

[58] Sheng-Hsiou Hsu, Tim Mullen, Tzyy-Ping Jung, and Gert Cauwenberghs. Online recursive independent component analysis for real-time source separation of high-density eeg. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, pages 3845–3848. IEEE, 2014.

[59] Ming-Zher Poh, Nicholas C Swenson, and Rosalind W Picard. Motion-tolerant magnetic earring sensor and wireless earpiece for wearable photoplethysmography. *IEEE Transactions on Information Technology in Biomedicine*, 14(3):786–794, 2010.

[60] Bernard Widrow, John R Glover, John M McCool, John Kaunitz, Charles S Williams, Robert H Hearn, James R Zeidler, JR Eugene Dong, and Robert C Goodlin. Adaptive noise cancelling: Principles and applications. *Proceedings of the IEEE*, 63(12):1692–1716, 1975.

[61] James AC Patterson, Douglas C McIlwraith, and Guang-Zhong Yang. A flexible, low noise reflective ppg sensor platform for ear-worn heart rate monitoring. In *Wearable and Implantable Body Sensor Networks, 2009. BSN 2009. Sixth International Workshop on*, pages 286–291. IEEE, 2009.

[62] David Da He. *A wearable heart monitor at the ear using ballistocardiogram (BCG) and electrocardiogram (ECG) with a nanowatt ECG heartbeat detection circuit*. PhD thesis, Massachusetts Institute of Technology, 2013.

[63] Moritz Kassner, William Patera, and Andreas Bulling. Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing: Adjunct publication*, pages 1151–1160. ACM, 2014.

[64] Paolo Gabriel, Werner K Doyle, Orrin Devinsky, Daniel Friedman, Thomas Thesen, and Vikash Gilja. 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*, pages 3402–3405. IEEE, 2016.

[65] Nancy XR Wang, Jared D Olson, Jeffrey G Ojemann, Rajesh PN Rao, and Bingni W Brunton. Unsupervised decoding of long-term, naturalistic human neural recordings with automated video and audio annotations. *Frontiers in human neuroscience*, 10, 2016.

[66] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 779–788, 2016.

[67] James D Morgante, Rahman Zolfaghari, and Scott P Johnson. A critical test of temporal and spatial accuracy of the tobii t60xl eye tracker. *Infancy*, 17(1):9–32, 2012.

[68] Dan Witzner Hansen and Qiang Ji. In the eye of the beholder: A survey of models for eyes and gaze. *IEEE transactions on pattern analysis and machine intelligence*, 32(3):478–500, 2010.

[69] Quoc V Le, Marc'Aurelio Ranzato, Rajat Monga, Matthieu Devin, Kai Chen, Greg S Corrado, Jeff Dean, and Andrew Y Ng. Building high-level features using large scale unsupervised learning. *arXiv preprint arXiv:1112.6209*, 2011.

[70] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-kin Wong, and Wang-chun Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. In *Advances in Neural Information Processing Systems*, pages 802–810, 2015.

[71] Yu Zhang, William Chan, and Navdeep Jaitly. Very deep convolutional networks for end-to-end speech recognition. *arXiv preprint arXiv:1610.03022*, 2016.

[72] Felix A Gers, Jürgen Schmidhuber, and Fred Cummins. Learning to forget: Continual prediction with lstm. *Neural computation*, 12(10):2451–2471, 2000.

[73] Rafal Jozefowicz, Wojciech Zaremba, and Ilya Sutskever. An empirical exploration of recurrent network architectures. In *Proceedings of the 32nd International Conference on Machine Learning (ICML-15)*, pages 2342–2350, 2015.