UNIVERSITY OF CALIFORNIA, SAN DIEGO

Finding Structure in Electrocorticographic Neural Signals for Brain-Machine Interface Applications

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Electrical Engineering
(Intelligent Systems, Robotics, and Control)

by

Akinyinka Oreoluwa Omigbodun

Committee in charge:

Professor Vikash Gilja, Chair
Professor Virginia de Sa
Professor Kenneth Kreutz-Delgado
Professor Bhaskar D. Rao
Professor Bradley Voytek

2017

The dissertation of Akinyinka Oreoluwa Omigbodun is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2017

DEDICATION

To God, and my Savior, Jesus Christ

EPIGRAPH

Swiftly the brain becomes an enchanted loom, where millions of flashing shuttles weave a dissolving pattern, always a meaningful pattern though never an abiding one; a shifting harmony of sub-patterns.

—Charles Sherrington

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I am very grateful for the outstanding mentorship that I have received from my advisor, Vikash Gilja, during the past three years. He has shown me effective strategies for balancing risks and rewards in pursuing research projects. Additionally, he has passed on a very handy "hourglass" approach to communicating about research: starting with the big picture, focusing on important details, and returning to the grand view. Thirdly, he has prompted very timely and constructive conversations about life and work in the course my doctoral studies. I know that I still have much to learn in these matters, but I am indebted to him for the foundation that he has laid.

I am thankful for the support that I have received from my committee members, Bradley Voytek, Virginia de Sa, Bhaskar Rao, and Ken Kreutz-Delgado, in words of encouragement and guidance; recommendation letters; and many other ways. I am deeply appreciative of the significant roles that Ebonee Williams, Tara Javidi, and Alon Orlitsky played in helping me navigate a very challenging transition during my doctoral research. I would like to thank Robert Lugannani for the opportunity to work under him as a teaching assistant on multiple occasions.

I have benefited greatly from many positive interactions with Terrance Mayes, Shana Slebioda, Adolfo Juarez, Samantha Garcia, Mary Duarte, Jennifer Neri, Chris Shenton, Kacy Vega, Zachary Dake, and other current and former advisors at the Electrical and Computer Engineering Graduate Student Affairs Office; I thank each of them. I also acknowledge Ricky Paniagua and other immensely helpful advisors at the International Students & Programs Office.

I thank Werner Jiang, Paolo Gabriel, Nick Rogers, John Hermiz, Tejaswy Pailla, Anthony Au, Francis Baek, Hannah Chou, Aashish Patel, Venkatesh Elango, Wahab Alasfoor, Kenny Chen, Daril Brown, other current and former colleagues of the Translational Neuroengineeing Laboratory, and Julie Moritz for a very stimulating and friendly work

environment. I also thank Gary Cottrell, Tomoki Tsuchida, Chris Kanan, Vicente Malave, Ben Cipollini, other members of the Cottrell Laboratory, and David Bareno for significant contributions to my development as a researcher.

I have been touched by the outpouring of support from Ananda Theertha Suresh, Yawo Ezunkpe, Menglin Zeng, Akshay Balsubramani, Anthony Nwokafor, and other current and former graduate students; as well as friends I have met through the UCSD Graduate Christian Fellowship, and Harbor Church. I must thank Thanos Siapas, Shuki Bruck, Azita Emami, and Chris Umans—at my undergraduate alma mater—who were instrumental in my journey to graduate school.

I thank Oluwadayo & Bolanle Oluwadara for their unrelenting interest in my progress since I started with higher education eleven years ago. I am very thankful for prayers and encouragement from dear friends and family across the globe.

Finally, I thank my parents, Akinyinka & Olayinka Omigbodun, and my sister, Iyeyinka. They have given me critical support every step of the way. Their constant encouragement, "Keep pushing!", has been golden.

Chapter 2, in part, is a reprint of the material as it appears in Omigbodun, A., Doyle, W. K., Devinsky, O., Friedman, D., Thesen, T., Gilja, V., "Hidden-Markov Factor Analysis as a Spatiotemporal Model for Electrocorticography", *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, IEEE, 2016. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in part, is currently being prepared for submission for publication of the material. Omigbodun, A., Doyle, W. K., Devinsky, O., Friedman, D., Melloni, L., Thesen, T., Gilja, V. The dissertation author was the primary investigator and author of this material.

Chapter 4, in part, has been submitted for publication of the material as it may

appear in Advances in Neural Information Processing Systems, 2017, Omigbodun, A., Doyle, W. K., Devinsky, O., Friedman, D., Melloni, L., Thesen, T., Gilja, V. The dissertation author was the primary investigator and author of this paper.

VITA

2010	B. S. in Electrical Engineering, California Institute of Technology
2013	Master of Science in Electrical Engineering (Intelligent Systems, Robotics, and Control), University of California, San Diego
2015	Candidate in Philosophy in Electrical Engineering (Intelligent Systems, Robotics, and Control), University of California, San Diego
2017	Doctor of Philosophy in Electrical Engineering (Intelligent Systems, Robotics, and Control), University of California, San Diego

PUBLICATIONS

Omigbodun, A., Doyle, W. K., Devinsky, O., Friedman, D., Thesen, T., Gilja, V., "Hidden-Markov Factor Analysis as a Spatiotemporal Model for Electrocorticography", *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, IEEE, 2016.

ABSTRACT OF THE DISSERTATION

Finding Structure in Electrocorticographic Neural Signals for Brain-Machine Interface Applications

by

Akinyinka Oreoluwa Omigbodun

Doctor of Philosophy in Electrical Engineering
(Intelligent Systems, Robotics, and Control)

University of California, San Diego, 2017

Professor Vikash Gilja, Chair

Electrocorticography (ECoG), also known as intracranial electroencephalography (iEEG), is the practice of recording electrical potentials on the cerebral cortex via electrodes placed on the exposed brain surface. ECoG has been a critical component of epilepsy medical treatment protocols involving neurosurgery for more than half a century. More recently, ECoG has emerged as a promising recording modality for brain-machine interfaces and neuroscience research. The BRAIN Initiative is representative of a renewed and concerted effort to push the boundaries of possibility in medical care and technology,

and to expand our understanding of brain function. Concomitant with this new drive is a need for techniques that address the challenges posed by high-channel count ECoG signal analysis as well as by neural data collection limited due to the invasiveness of ECoG. In this dissertation, we introduce the use of a discrete-state based probabilistic method for modeling ECoG-derived signals, and contrast this method with previously existing analogous probabilistic models without a discrete component. We then explore a class of discrete-state based probabilistic models, and identify spatial and temporal model constraints that were advantageous in the analysis of a high-channel count ECoG dataset. Finally, we introduce another probabilistic model that we use for unsupervised learning of ECoG trial spatiotemporal structure, and clustering of the ECoG trials in a data-limited context.

Chapter 1

Introduction

Electrocorticography (ECoG) is a neural recording modality involving the intracranial placement of electrodes on the cerebral cortex for electrical potential measurements. ECoG lies at the center of a plexus of applications. More than half a century ago, Wilder Penfield and Herbert Jasper pioneered the use of ECoG as part of the protocol in the surgical treatment of intractable epilepsy. They used ECoG to identify cortical regions important for critical functions like speech processing, while excising noncritical brain tissue implicated in epilepsy. Penfield also made important contributions to Neuroscience research, providing fairly detailed maps associating regions on the surface of the cortex with the sensory stimulation and motor activity of specific anatomical divisions of the body. Today, neuroscientists continue to employ ECoG in probing the mechanisms underlying different aspects of perception, action, and cognition.

A more recent application of ECoG is in the area of brain-machine interfaces (BMIs). In 2013, Wang *et al*. published results of an ECoG BMI research study with a tetraplegic subject [WCD$^+$13], demonstrating the control of a display cursor through a neural link (communication neuroprosthetic) and the use of a robotic arm to hit targets (motor neuroprosthetic). Individuals affected by tetraplegia are only one segment among

others in the population that would benefit from versions of the BMIs in this study that are commercially viable and robust under a variety of conditions and across individuals.

One program of research that may be crucial to advancing to this point of widespread BMI accessibility is identifying models of neural activity that allow us to more clearly express the relationship between perception, behavior, and cognition and properties of the ECoG signal of groups of electrodes in a multi-electrode recording—properties such as average activity, shared variability, and the neuroanatomical placement of the electrodes.

An emerging approach to modeling neural activity is to learn a function or set of functions that maps the activity at one point in the ECoG state space to the next point in time. This is the approach exemplified by dynamical systems and recurrent neural networks. The nature of the ECoG signal is such that the kinds of functions that would be appropriate with these approaches would be highly nonlinear. Indeed, nonlinear dynamical systems and neural networks are designed just for that. However, the invasiveness of the ECoG recording modality means that we are often data-limited which could be an obstacle to learning such functions. High-dimensionality is also a challenge as we are dealing with tens to hundreds of electrodes with only tens of minutes of ECoG recordings.

The general approach that we take here is one of clustering or discretization in the ECoG state space, and the ultimate goal is to build rigorous connections between the ECoG state space with the clustering and the much more abstract "neural state" space of perception, behavior and cognition. It is our hope that the work in this dissertation represents at least a few steps in the right direction.

In Chapter 2, we juxtapose our probabilistic approach to quantizing the neural space with other probabilistic approaches that are closer to the dynamical systems approach. Our method makes use of signal variance or noise to define neural activity

transitions in state space while the other techniques use temporal smoothing to reduce noise in the neural trajectories in state space. In Chapter 3, we explore a class of discrete-state based probabilistic models, and identify spatial and temporal model constraints that were advantageous in the analysis of a high-channel count ECoG dataset. The various spatial constraints entail different assumptions about the ECoG signal covariance. We consider models with and without a Markov temporal constraint. In Chapter 4, we present a probabilistic model that we developed for the unsupervised learning of ECoG trial spatiotemporal structure, and clustering of the ECoG trials in a data-limited context. The models in our treatment here are extensions of models advanced in Chapter 3. Chapter 5 contains concluding remarks and proposed future directions.

Chapter 2

Hidden-Markov Factor Analysis as a Spatiotemporal Model for Electrocorticography

We present a new approach to extracting low-dimensional neural trajectories that summarize the electrocorticographic (ECoG) signals recorded with high-channel-count electrode arrays implanted subdurally. In our approach, Hidden-Markov Factor Analysis (HMFA), a finite set of factor analyzers are used to model the relationship between the high-dimensional ECoG neural space and a low-dimensional latent neural space; the factor analyzers at different time points are in turn linked together with a hidden Markov model. The recorded ECoG signals were band-pass filtered such that our analysis was focused on a sub-band (76-100Hz) of high gamma. HMFA affords the quantization of the ECoG neural space and dimensionality reduction in a common probabilistic space. We applied this method to the ECoG recordings of two subjects who responded with button presses to audiovisual stimuli in an experimental task. Our results from comparing HMFA with Gaussian-Process Factor Analysis (GPFA) and other related spatiotemporal

modeling techniques, using a goodness-of-fit metric that measures how well the ECoG activity of each electrode can be reconstructed from the other electrodes, suggest that switching between multiple linear models may better capture neural activity across the electrodes. In contradistinction to HMFA, GPFA and the other techniques integrate temporal smoothing with a single linear dimensionality reduction model. We believe that HMFA will provide a powerful tool for relating high-channel-count ECoG signals to the perception and behavior of subjects.

2.1 Introduction

With the advent of the BRAIN Initiative, we are closer to an era in which cortical implants with hundreds and even thousands of neural probes are standard in brain diagnostic and brain-machine interface (BMI) applications. While such technology could help with unlocking new perspectives about cortical activity hitherto unavailable, the challenge of generating readily interpretable views of cortical dynamics will become even greater. In cases where cortical mechanisms are observable only by considering collections of channels recording neural activity and not at the level of single electrodes [CY14], the statistical challenges associated with with using experimental trial averages to combat noise increase. Moreover, subtle aspects (e.g. slight timing differences, etc.) of perception, decision making, attention, or motor planning can be obscured altogether with approaches primarily based on trial averages. There is a greater urgency for approaches that facilitate single-trial analyses [YCS$^+$09, ASGA96]. With single-trial analytical methods, experimental conditions may not have to be as tightly controlled. In the context of analyzing neural spiking data, Yu *et al.* developed one such approach, GPFA, in which they combined temporal smoothing, to account for noise, with dimensionality reduction, to generate lower dimensional neural signal representations that can be more readily

interpreted.

The primary motivation for our work was to consider the related problem of simultaneously extracting low-dimensional neural trajectories and identifying a finite set of low-dimensional subspaces that contain the extracted neural trajectories. The manner in which these neural trajectories — for different experimental epochs — move between these subspaces could be used to relate the epochs in terms of human perception and behavior in an unsupervised fashion [WOO$^+$16]. Figure 2.1 illustrates how a neural trajectory is extracted from a multi-electrode recording using HMFA. The extracted high gamma power from a hypothetical 2-electrode recording is shown in panel A of Figure 2.1. We have chosen to have 2 electrodes in the illustration for simple comprehension of the problem we approach in the high-dimensional space (typically with tens to hundreds of electrodes). In panel B, a high-dimensional (2-dimensional) space is shown where each axis corresponds to the activity of an electrode. Shown in panel C are 3 low-dimensional (1-dimensional) subspaces learned in our approach. The number of subspaces to be learned is specified for the model. In panel D, ellipses corresponding to the subspaces are shown to indicate a quantization of the high-dimensional space. We note here that because of the temporal structure of the model, the proximity of a point along a trajectory to the center of an ellipse does not necessitate its projection into the corresponding subspace. Finally, the neural trajectories extracted are in panel E. They are constrained to lie in the subspaces. We can view each neural trajectory as a progression among the subspaces.

The key distinguishing features of HMFA from GPFA and the other techniques discussed here are (1) the absence of temporal smoothing and (2) presence of multiple low-dimensional subspaces. We first describe HMFA—an approach that enables us to integrate the quantization and dimensionality reduction operations on the neural space. We also adapt a goodness-of-fit metric [YCS$^+$09] that allows us to compare HMFA with

GPFA and other methods, in terms of the ability to reconstruct electrode channel activity. With these in place, we apply HMFA and the other neural trajectory extraction methods to ECoG recordings from a cross-modal speech experiment.

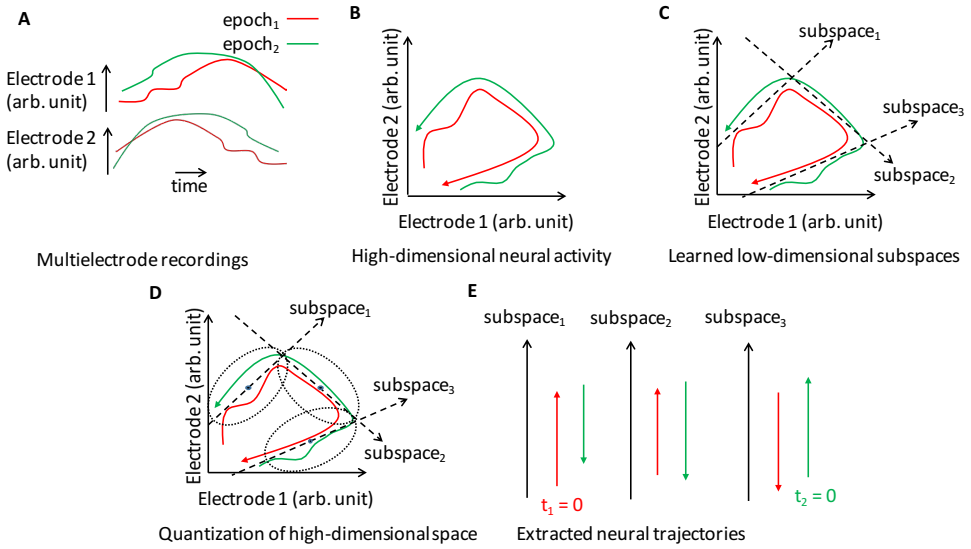


Figure 2.1: Extracting neural trajectories from multielectrode recordings with Hidden-Markov Factor Analysis. For clarity the activity of only 2 electrodes is considered in this illustration. A: activity recorded simultaneously from 2 electrodes (after signal processing). B: the time evolution of the recorded neural activity plotted in a 2-dimensional space, where each axis corresponds to the activity of an electrode. C: 3 learned 1-dimensional subspaces. D: quantization of high-dimensional space E: neural trajectories viewed in the low-dimensional subspaces

2.2 Methods

2.2.1 Hidden-Markov Factor Analysis

In Hidden-Markov Factor Analysis (HMFA), a finite set of factor analyzers are used to model the relationship between the high-dimensional neural space and a low-dimensional latent neural space; factor analyzers at different time points are linked together with a hidden Markov model (HMM). While HMFA has been used in other domains for clustering in gene expression analysis [MM12] and classification of acoustic features in speech recognition [SR98], we are—to the best of our knowledge—the first to pursue the unsupervised analysis of ECoG neural signals with it.

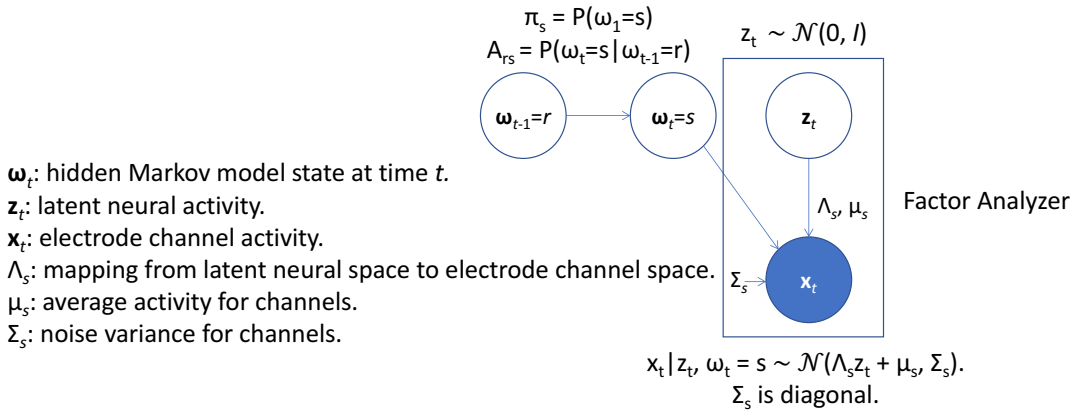


Figure 2.2: Hidden-Markov Factor Analysis

Let $\mathbf{x}_t$ of dimension n denote the high-dimensional vector of the preprocessed electrode channels' activity at time point $t \in \{1,...,T\}$, where n is the number of electrodes. Our goal is to extract a corresponding low-dimensional latent neural state $\mathbf{z}_t$ of dimension $l < n$, as well as the index $\omega_t \in \{1,...,S\}$ of the factor analyzer at that time point. We define the following linear-Gaussian relationship between the electrode activity $\mathbf{x}_t$ and the latent neural state $\mathbf{z}_t$:

$$\mathbf{x}_t|\mathbf{z}_t, \omega_t = s \sim \mathcal{N}(\Lambda_s \mathbf{z}_t + \mu_s, \Sigma_s), \tag{2.1}$$

where Λ_s (n by l matrix), μ_s (n-dimensional vector), and Σ_s (n by n matrix) are model parameters to be learned for $s \in \{1,...,S\}$. Λ_s is referred to as a factor loading matrix. As is standard in Factor Analysis (FA), we constrain the covariance matrix Σ_s to be diagonal, where the diagonal elements are the independent noise variances of each electrode. We also define $\mathbf{z}_t$ as a multivariate standard normal random variable, which is typical in FA (2.2).

$$\mathbf{z}_t \sim \mathcal{N}(0, I). \tag{2.2}$$

The factor analyzers at different time points are related through an HMM to capture the idea of neural trajectories moving within as well as between low-dimensional subspaces. Start and transition probability parameters are learned for the S-state HMM.

$$\pi_s = P(\omega_1 = s),\ \ A_{rs} = P(\omega_t = s|\omega_{t-1} = r),\ \ s,r \in \{1,...,S\}. \tag{2.3}$$

The parameters of the HMFA model $\Theta == \{\pi, A, (\mu_s, \Lambda_s, \Sigma_s)_{s=1}^S\}$ can be fit using a variant of the widely used expectation maximization (EM) algorithm known as the Alternating Expectation Conditional Maximization (AECM) algorithm [MM12, MVD97]. The algorithm seeks the model parameters that maximize the probability of the observations: preprocessed electrode activity in this case. The parameter S must be specified before the model fitting can be done.

Once the HMFA model is learned, we can use the Viterbi algorithm [Rab89] to compute the most probable sequence of factor analyzer states $\Omega_{\text{Viterbi}} = [\omega_1, ..., \omega_T]$. For notational convenience, let us group the latent neural states for different time points together in an l by T matrix, $\mathbf{Z} = [\mathbf{z}_1, ..., \mathbf{z}_T]$ and the observations in an n by T matrix, $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_T]$. With the Viterbi state sequence path Ω_{Viterbi}, we can extract neural trajectories $\mathbb{E}[\mathbf{Z}|\mathbf{X}, \Omega_{\text{Viterbi}}]$ from the preprocessed electrode activity $\mathbf{X}$ by taking advantage of the jointly Gaussian relationship between $\mathbf{Z}$ and $\mathbf{X}$ given Ω_{Viterbi}.

2.2.2 Leave-one-electrode-out Reconstruction Error

In order to compare HMFA with GPFA and other methods considered in [YCS+09], we adapted a goodness-of-fit metric that Yu *et al.* introduced in their work. For each of the methods, model parameters were fit to training data. With data not used for model fitting, we leave out each electrode in turn and compute how well the fitted model can be used to reconstruct the activity of an electrode, given the activity of all the other electrodes.

Let $\mathbf{X}_{-j}$, (n-1) by T matrix, be the matrix of observations without the j^{th} electrode and $\mathbf{X}_j$, of dimension T, be the vector of the j^{th} electrode observations. Once again taking advantage of a jointly Gaussian relationship—this time—between $\mathbf{X}_j$ and $\mathbf{X}_{-j}$ given Ω_{Viterbi}, we can use $\mathbb{E}[\mathbf{X}_j|\mathbf{X}_{-j}, \Omega_{\text{Viterbi}}]$ as our reconstruction of the j^{th} electrode observations. (Please see **Leave-one-electrode-out Prediction Error** in Subsection 3.2.4.) The reconstruction error (*RE*) is defined as the sum-of-squared differences between the model reconstruction and the observed activity across all electrodes and time points.

$$RE = \sum_{j=1}^{n} ||\mathbb{E}[\mathbf{X}_j|\mathbf{X}_{-j}, \Omega_{\text{Viterbi}}] - \mathbf{X}_j||. \tag{2.4}$$

2.2.3 Cross-modal Speech Experiment and ECoG Recordings

Experimental protocols were approved by the New York University School of Medicine Institutional Review Board and subjects undergoing ECoG-based cortical mapping for brain surgery gave their informed consent to participate in the experiment. In the experiment, the subjects were presented with a series of audiovisual stimuli involving a speaker on a screen. The subjects were instructed to press a button whenever they heard or saw the speaker say one of two words: “cafe” and “avenue”. In terms of the audiovisual stimuli, there were 7 trial types in the experiment: (1) stimuli for which

the word heard matched the word seen (2) audio-only stimuli (3) stimuli for which the word heard did not match the word seen (4) video-only stimuli (5) video with the speaker making grotesque mouth movements (6) phase-scrambled audio with the speaker making grotesque mouth movements (7) phase-scrambled audio alone. Valid button presses were only possible in trial types (1), (2) and (4); indeed, none of the subjects registered button responses in any of the other trial types. In summary, 10 trial types were possible in terms of the stimuli and behavioral response. The experiment was divided into 5 blocks of 82 trials.

We used the data of two subjects, NY451 and NY453, in our analyses here. Electrodes in both grid and strip configurations were implanted subdurally. We used only the grid electrodes in our work. There were 112 electrodes for each subject: 64 in an 8 by 8 grid with smaller electrode separation and 48 in another 8 by 8 grid with larger separation. (The grids overlapped with the smaller grid taking the place of 16 electrodes in the larger grid.) The grid electrodes were implanted over the right hemisphere for both subjects. The ECoG signal was sampled at 512 Hz and measured in microvolts.

We carried out the following signal processing steps: (1) Electrodes with unusual—more than 3 standard deviations from the mean across the electrodes of—rms power, maximum or median absolute potentials, or mean potentials were rejected from our analyses. (2) We used notch filtering to remove the harmonics of line noise at 60 Hz, 120 Hz, 180 Hz and 240 Hz respectively. (3) We applied common average referencing to eliminate noise common to all the ECoG channels. (4) We removed the best fit line in the least-squares sense from each ECoG channel time series to eliminate systematic overall increases and decreases in electrode potentials. (5) We band-pass filtered the 76-100 Hz sub-band of high gamma. Thereby, we attempted to limit the effects of the 1st and 2nd harmonics of line noise on our analyses. There are a number of studies that have shown connections between gamma-band activity and motor behavior [MLS$^+$07] and

perceptual cognition [JK09]. (6) We extracted 5-second windows of trials from the ECoG recordings: 2 seconds before the stimulus onset to 3 seconds after. (7) In a similar fashion to the electrode rejection, we eliminated trials with unusual ECoG activity. (8) Lastly, with a 50-millisecond sliding window we computed the rms power in nonoverlapping bins for each trial. We analyzed a total of 236 trials (from 3 blocks) for NY451 and 245 trials (from 3 blocks) for NY453, with 105 and 109 electrodes respectively.

2.3 Results

Using the goodness-of-fit metric described in Subsection 2.2.2, we compared HMFA, GPFA, reduced GPFA, and approaches in which the preprocessed electrode activity was first smoothed over time using a Gaussian kernel before the application of static dimensionality-reduction techniques: Principal Components Analysis (PCA), Probabilistic PCA (PPCA), or Factor Analysis (FA) [YCS$^+$09]. We shall refer to the latter as two-stage methods hereinafter. In GPFA, the columns of the estimated factor loading matrix correspond to directions in the low-dimensional space. While not orthogonal in general, the columns can be orthonormalized and ordered by the amount of data covariance explained. The top $l'(l' \leq l)$ dimensions can be used to extract an l'-dimensional neural trajectory for the reduced GPFA model. Please note that when l' and l are equal, GPFA and reduced GPFA are identical. We also note that HMFA with 1 hidden state is equivalent to FA without temporal smoothing. We used 4-fold cross-validation in all the analyses.

Figure 2.3, panel A(I) shows the reconstruction error for NY451 for all the methods across different latent state dimensionalities (l = 1–25) with a Gaussian kernel width of 0.14 seconds for the two-stage methods and 4 hidden states for HMFA. (The reconstruction error for HMFA stops at l = 8 due to data limitations on the parameter es-

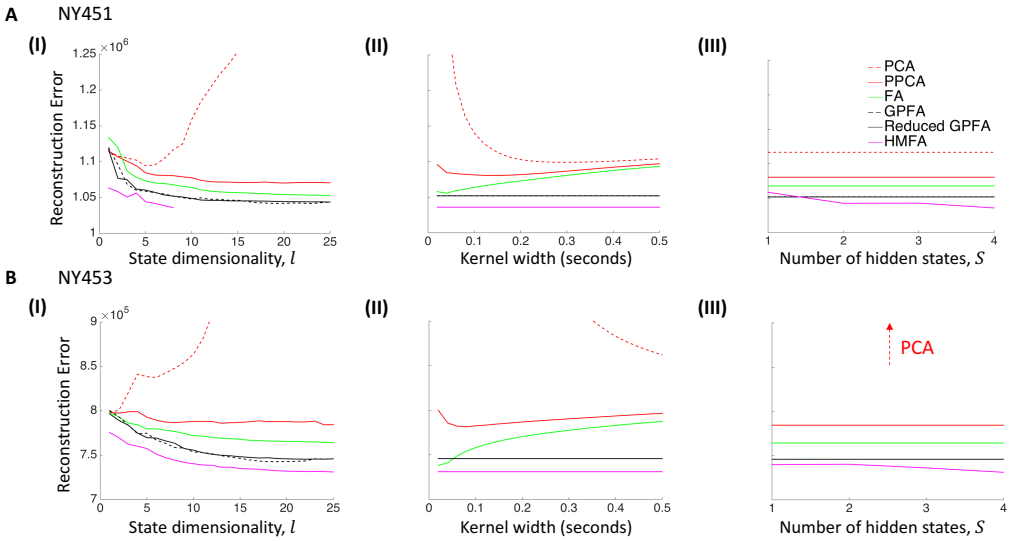


Figure 2.3: Reconstruction error. A: For NY451 across (I) state dimensionalities (kernel width = 0.14 seconds and $S = 4$). (II) kernel widths ($l = 8$ and $S = 4$). (III) numbers of hidden states (kernel width = 0.14 seconds and $l = 8$). B: For NY453 across (I) state dimensionalities (kernel width = 0.14 seconds and $S = 4$). (II) kernel widths ($l = 25$ and $S = 4$). (III) numbers of hidden states (kernel width = 0.14 seconds and $l = 25$).

timation.) Figure 2.3, panels A(II)-(III) show the reconstruction error for all the methods across Gaussian kernel widths 0.02–0.5 seconds and 1–4 hidden states respectively. For $l \geq 3$, PCA, PPCA, and FA resulted in lower reconstruction errors respectively (Wilcoxon paired-sample test, $P < 0.001$). Yu *et al* observed a similar trend in their analyses of neuronal recordings. Statistical significance was assessed by looking across all trials for each method at its optimal latent state dimensionality l, Gaussian kernel width, and/or number of hidden states, S. Contrary to observations by Yu *et al*, we did not observe lower reconstruction error with GPFA relative to the two-stage FA method, or with reduced GPFA relative to GPFA. However, while their analysis [YCS$^+$09] was done on neural spiking data, we used ECoG data in our work presented here. HMFA produced the lowest reconstruction error among all the methods (Wilcoxon paired-sample test, $P < 0.001$). Figure 2.3, panels B(I)-(III) show reconstruction error with different methods for

subject NY453, and HMFA also produced the lowest reconstruction error among all the methods (Wilcoxon paired-sample test, $P < 0.001$).

2.4 Discussion

We have introduced a new method, HMFA, for extracting single-trial neural trajectories from ECoG recordings. We have also adapted a goodness-of-fit metric that has allowed us to compare the ability to reconstruct channel activity of HMFA with GPFA, and other neural trajectory extraction methods. While temporal smoothing is not incorporated in HMFA as it is in the other methods, only HMFA allows for multiple low-dimensional subspaces. We have explored the space of HMFA models with 1–4 low-dimensional subspaces (hidden states). With 4 hidden states, the reconstruction error of HMFA was lower than those of the other methods with the data from both subjects. In the following chapters, we will explore the relationship between HMFA-derived state sequences, and the perception and behavior of subjects.

Acknowledgement

Chapter 2, in part, is a reprint of the material as it appears in Omigbodun, A., Doyle, W. K., Devinsky, O., Friedman, D., Thesen, T., Gilja, V., "Hidden-Markov Factor Analysis as a Spatiotemporal Model for Electrocorticography", *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, IEEE, 2016. The dissertation author was the primary investigator and author of this paper.

Chapter 3

Discrete state based approaches to analyzing electrocorticographic data

We explore a class of latent variable model–based clustering methods in the context of high-channel count electrocorticographic (ECoG) signal analysis. The ECoG signals were recorded with electrode arrays implanted subdurally. Our analysis was focused on a sub-band (76-100 Hz) of high gamma. In each of the methods, the pre-processed ECoG neural activity at each time point was modeled with a mixture of multivariate normal distributions, which accounts for shifts in average neural activity as well as changes in channel covariance structure over time. We applied the methods to the ECoG recordings of six subjects who responded with button presses to audiovisual stimuli in an experimental task. We used two metrics in comparing the methods: (1) a goodness-of-fit metric that measures how well the ECoG activity of each electrode can be predicted by all the other electrodes and (2) the mutual information between the ECoG experimental task trial label and the most probable mixture component at each time point. The four key features of interest in the comparison of models were (1) the ability to model channels' shared variance, (2) a constraint on how much shared variance

is captured (by models with the ability to model shared variance), (3) invariance across mixture components in a channel private noise or variance parameter (for models with the shared variance constraint), and (4) a Markov temporal constraint. We found that models with all four features were the most successful at encoding stimuli and behavior information of the trials as well as modeling changes in channel covariance structure.

3.1 Introduction

As medical doctors, neuroscientists, and neuroengineers push towards more effective prevention and treatment of brain dysfunction, greater understanding of brain structure and function, and more successful brain-machine interface (BMI) system design, it is likely that the use of cortical implants with hundreds and even thousands of neural probes will become more widespread. One of the challenges that this trend will present is generating readily interpretable perspectives of high channel count electrode signals. In situations where cortical mechanisms can be observed only by considering multi-channel neural activity recordings and not at the level of single electrodes [CY14], greater statistical challenges are associated with using experimental trial averages to deal with noise. In addition, subtle aspects (e.g. slight timing differences, etc.) of decision making or motor planning may be lost in using approaches primarily based on trial averaging. Moreover, many methods—single-trial and otherwise—have the assumption that the covariance structure of the neural activity channels is constant in the course of an experimentally defined trial [BMJC13, YCS$^+$09], which may not necessarily be the case [CKC$^+$11].

Hidden-Markov Factor Analysis (HMFA) and the other methods that we explore here allow us to account for changes in covariance structure over the duration of an experimental trial. We consider the benefits of different spatial and temporal constraints

specific to different methods. We identify a class of HMFA models that performs better than other models in accounting for different stimuli and behavioral responses along with associated ECoG neural signal means and covariances in a cross-modal speech experiment.

3.2 Materials and Methods

3.2.1 Cross-modal Speech Experiment and ECoG Recordings

The description of the experiment, recordings, and signal processing is abbreviated here. Please refer to a more complete description in Subsection 2.2.3. The number of subjects in the analyses here was expanded to six from two. Tables 3.1 and 3.2 contain information on the electrodes and trials respectively for each subject.

Electrodes in both grid and strip configurations were implanted subdurally. Again, we used only the grid electrodes in our work. There were 112 electrodes for NY441, NY451, NY453 and NY468: 64 in an 8 by 8 grid with smaller electrode separation and 48 in another 8 by 8 grid with larger separation. (The grids overlapped with the smaller grid taking the place of 16 electrodes in the larger grid.) NY400 and NY455 had 64 electrodes in a single 8 by 8 grid with larger separation. We note that the activity for a number of the electrodes was not recorded with the signal acquisition systems for NY400 and NY441 and hence, there is a discrepancy between the number of electrodes available before signal processing for NY400 and NY455 as well as between those of NY441 and NY451, NY453, and NY468.

There were modifications to some of our ECoG signal extraction[1,2] and process-

[1]We carried out our signal processing steps on the complete ECoG recording available rather than limiting signal processing to the portion of the recording particularly pertinent to the Cross-modal Speech Experiment.

[2]We also introduced an experimental block with incomplete data for NY453. (NY451, NY453, NY455, and NY468 had blocks that were missing and/or incomplete.)

ing[3,4] steps detailed in Subsection 2.2.3 that resulted in different numbers of electrodes and trials than before for NY451 and NY453.

Table 3.1: Cross-modal speech subject electrode information

Subject	Number of electrodes (grid only)		Hemisphere implanted with grid(s)
	Before signal processing	After signal processing	
NY400	32	31	Left
NY441	104	99	Left
NY451	112	109	Right
NY453	112	107	Right
NY455	64	58	Left
NY468	112	105	Right

Table 3.2: Cross-modal speech subject trial information

Subject	Number of valid experimental blocks	Number of valid trials	
		Before signal processing	After signal processing
NY400	5	410	410
NY441	5	410	410
NY451	3	238	236
NY453	4	288	287
NY455	5	409	409
NY468	5	406	406

[3]There were two data acquisition systems used in parallel to record from the neural implants for all the subjects except for NY400. Here, we limited common average referencing to channels belonging to the same system.

[4]In addition to a neural signal time step of 50 milliseconds, we also considered 100 and 200 milliseconds in our analyses in Chapter 3.

3.2.2 Latent Variable Models

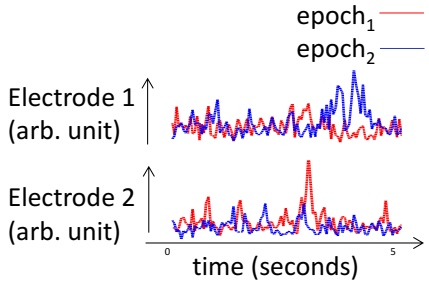
Gaussian Mixture Model with Full Covariance ($\mathbf{GMM_{full}}$)

GMM_{full} is a statistical model that relates a real-valued observed variable such as ECoG neural activity to a discrete latent variable. The discrete latent variable can take on the values of the members of a set of discrete values. Each discrete value in the set corresponds to a Gaussian distribution parameterized by a mean and full covariance. The Gaussian distributions are the mixture components of the model.

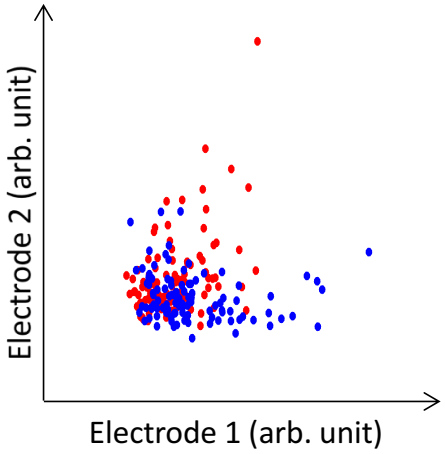
Figure 3.1a shows the ECoG activity (after signal processing) of 2 electrodes during 2 experimental trials for a subject. In Figure 3.1b, we have a state-space representation of the ECoG activity from the 2 trials, with each axis corresponding to the activity for an electrode. Many more trials are included in Figure 3.1c. In Figure 3.2a, a depiction of the 3 mixture components of a GMM_{full} model fit to the ECoG activity in Figure 3.1c is shown with the trial time points color-coded according to the most probable Gaussian mixture component at each time point. (The depiction is purely for illustrative purposes.)

Gaussian Mixture Model with Diagonal Covariance ($\mathbf{GMM_{diag}}$)

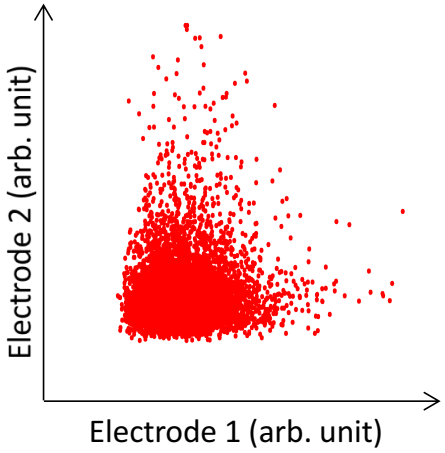
GMM_{diag} differs from GMM_{full} in that the covariance of each mixture component is diagonal. Only the variance specific to each channel is captured by the model. The reduction in model complexity from GMM_{full} to GMM_{diag} is significant when the number of electrode channels is large. In Figure 3.2b, a depiction of the 3 mixture components of a GMM_{diag} model fit to the ECoG activity in Figure 3.1c is shown. Please note that the covariance ellipse for each mixture component is aligned with the coordinate axes just as we should expect for GMM_{diag}.



(a) Time series representation (two trials)

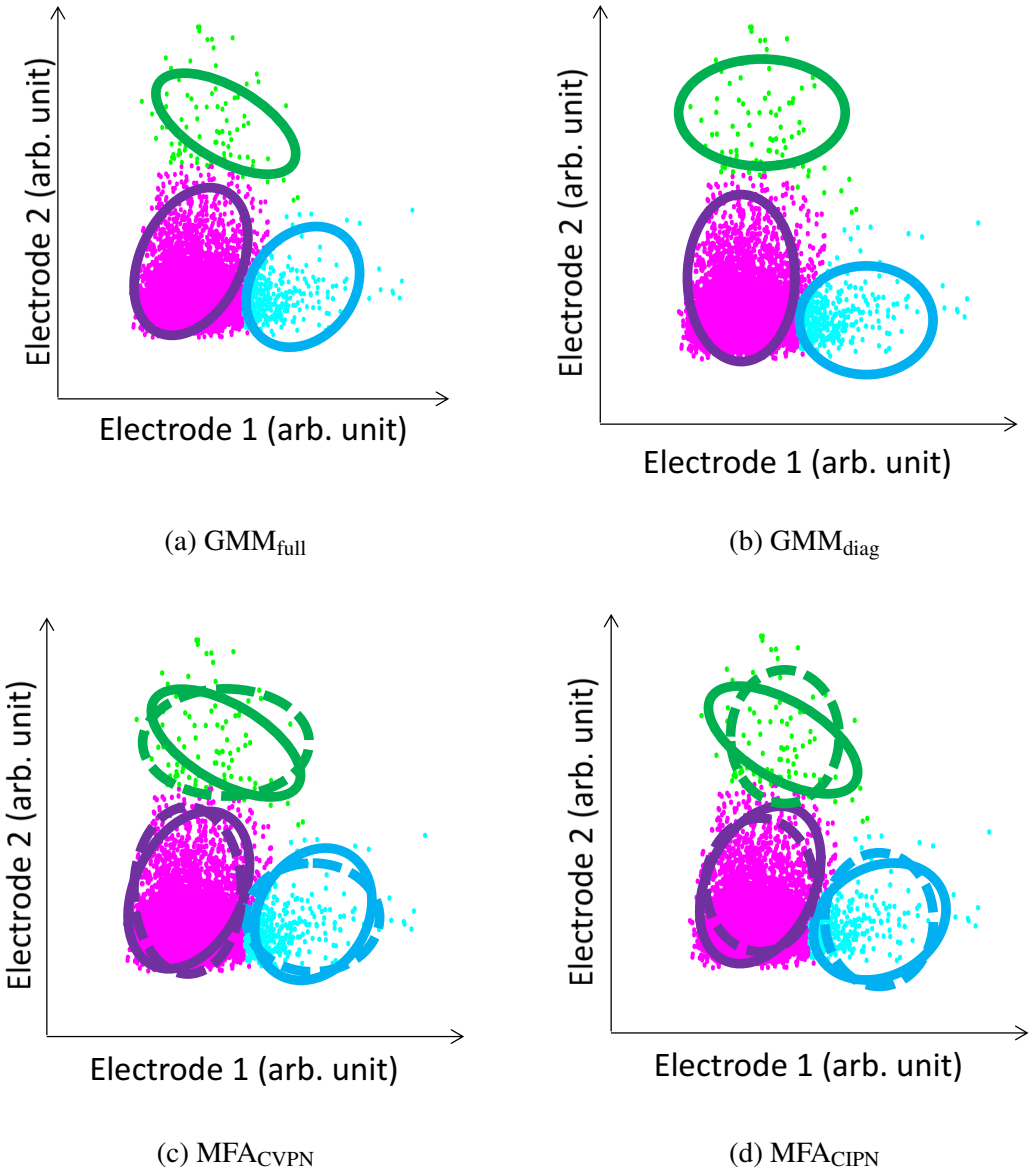


(b) State-space representation (two trials)



(c) State-space representation (many trials)

Figure 3.1: Time series and state space representations



(a) GMM_{full} (b) GMM_{diag}

(c) MFA_{CVPN} (d) MFA_{CIPN}

Figure 3.2: Latent variable models (various spatial constraints)

Mixture of Factor Analyzers with Component-Variant Private Noise (MFA_{CVPN})

MFA_{CVPN} possesses a real-valued latent variable in addition to a discrete latent variable. The discrete latent variable plays a similar role as in GMM_{full} and GMM_{diag}. However, the real-valued latent variable, known as a factor, can be viewed as a lower-dimensional latent neural state corresponding to the observed higher-dimensional ECoG neural activity. Let $\mathbf{x}_t$ of dimension n represent the ECoG neural activity at time t, where n is the number of electrodes; let $\mathbf{z}_t$ of dimension $l < n$ denote the factor at time t, and $\omega_t \in \{1, ..., S\}$, the discrete latent variable. We have the following linear-Gaussian relationship between $\mathbf{x}_t$ and $\mathbf{z}_t$, given ω_t:

$$P(\mathbf{x}_t|\mathbf{z}_t, \omega_t = s) = \mathcal{N}(\Lambda_s \mathbf{z}_t + \mu_s, \Sigma_s), \tag{3.1}$$

where Λ_s (n by l matrix), μ_s (n-dimensional vector), and Σ_s (n by n matrix) are model parameters to be learned for $s \in \{1, ..., S\}$. Λ_s is called a factor loading matrix. Σ_s is constrained to be diagonal and

$$P(\mathbf{z}_t) = \mathcal{N}(0, I). \tag{3.2}$$

It follows from Equations 3.1 and 3.2 that:

$$P(\mathbf{x}_t|\omega_t = s) = \mathcal{N}(\mu_s, \Lambda_s(\Lambda_s)' + \Sigma_s), \tag{3.3}$$

Consequently, MFA_{CVPN} is a model with Gaussian mixture component covariances of intermediate complexity between GMM_{full} and GMM_{diag}. Λ_s captures the shared variance or noise and Σ_s models the private noise. Each mixture component has its own private noise parameter and hence the private noise is component-variant. Finally, MFA_{CVPN} also has a parameter, π of dimension s, for the mixture component prior

probabilities:

$$\pi_s = P(\omega_t = s). \tag{3.4}$$

In Figure 3.2c, a depiction of the 3 mixture components of a MFA_{CVPN} model fit to the ECoG activity in Figure 3.1c is shown. The covariance ellipses for private noise are shown with dashed lines.

Mixture of Factor Analyzers with Component-Invariant Private Noise (MFA_{CIPN})

MFA_{CIPN} differs from MFA_{CVPN} in that the private noise is component-invariant; the mixture components have a single private noise parameter. Ghahramani & Hinton [GH96] derived an expectation maximization (EM) algorithm for learning the parameters of MFA_{CIPN}; a slight modification of the algorithm is applicable to MFA_{CVPN}. In Figure 3.2d, a depiction of the 3 mixture components of a MFA_{CIPN} model fit to the ECoG activity in Figure 3.1c is shown. The covariance ellipses for private noise, shown with dashed lines, all have the same shape. One motivation for considering MFA_{CIPN} together with MFA_{CVPN} was to explore the connection between ECoG changing private noise and stimulus-behavior information as has been done with single-neuron recordings [CKC$^+$11].

Hidden Markov Model (HMM) and Hidden Markov Factor Analysis (HMFA)

HMM (full or diagonal) and HMFA (CVPN or CIPN) have almost all the machinery of GMM (full or diagonal) and MFA (CVPN or CIPN) respectively but in lieu of a parameter for the mixture component prior probabilities, there are start and transition probability parameters:

$$\pi_s = P(\omega_1 = s),\ \ A_{rs} = P(\omega_t = s | \omega_{t-1} = r),\ \ s, r \in \{1, ..., S\}. \tag{3.5}$$

These enable a temporal component of modeling for the ECoG neural activity. The parameters of the models, $\Theta_{\text{HMM}} = \{\pi, A, (\mu_s, \Sigma_s)_{s=1}^{S}\}$ and $\Theta_{\text{HMFA}} = \{\pi, A, (\mu_s, \Lambda_s, \Sigma_s)_{s=1}^{S}\}$, can be learned with the Baum–Welch EM-based algorithm [Rab89] and a variant of the EM algorithm known as the Alternating Expectation Conditional Maximization (AECM) algorithm [MM12] respectively. Then, with the Viterbi algorithm we can infer the most probable sequence of mixture components, Ω_{Viterbi}. Figure 2.2 shows two time points and the factor analyzer at one of them in a graphical model of $\text{HMFA}_{\text{CVPN}}$.

Mixture Model augmented by a Markov Model (GMM-MM or MFA-MM)

We introduced the parameter estimation approach here to probe the advantages of joint spatial and temporal parameter estimation for HMM and HMFA. In the first stage, the parameters of a GMM (full or diagonal) or an MFA model (CVPN or CIPN) are learned, and then the latent variables are inferred given the model and ECoG data. In the second stage, the inferred discrete latent variables (specifically the indices of the most probable mixture components) are treated as observed variables and used in learning the parameters of a Markov model. GMM-MM and MFA-MM have the same parameters as HMM and HMFA respectively.

3.2.3 Model Fitting

We employed 4-fold cross-validation in model fitting. We ensured that all 10 trial types were present in each cross-validation fold. In addition, we limited the variation in the proportions of trial types among different folds. Otherwise, the assignment of trials to cross-validation folds was random. The learning tolerance threshold was 10^{-2} for all models with the exception of the GMM models[5], for which it was 10^{-6}.

[5]This was to account for the fact that the MATLAB Gaussian Mixture Model EM algorithm stopping criterion is implemented differently.

3.2.4 Metrics

Normalized Mutual Information (*NMI*)

For each latent variable model, we can infer the most probable mixture component at each time point for each trial, ω_t. We define the ground truth discrete state at each time point as the trial type index, ξ_t ($\xi_t \in \{1, ..., 10\}$). (The ideal choice of ground truth discrete state would be an index corresponding to the stimuli and behavior at each time point of a trial. However, our access to such information is limited.)

Then, we can compute the mutual information, $I(\Omega;\Xi)$. Intuitively, $I(\Omega;\Xi)$ is a measure of how much ω_t tells us about ξ_t or vice-versa. We note that because ξ_t is by definition constant for each trial, if we compute $I(\Omega;\Xi)$ limited to a single trial the result is 0; the same is true if the computation of $I(\Omega;\Xi)$ is limited to a single trial type. We endeavored in our analyses here to ensure that all trial types are present and the proportions of trial types is similar across all computations of $I(\Omega;\Xi)$. The computation of $I(\Omega;\Xi)$ was limited to each cross-validation fold due to the fact that there was no straightforward correspondence between mixture components learned for different cross-validation folds. Finally, $I(\Omega;\Xi)$ was normalized as shown in Equation 4.20 to obtain the normalized mutual information.

Leave-one-electrode-out Prediction Error (*PE*)

For each model, with data not used in model fitting, we leave out each electrode in turn and compute how well the model can predict the activity of an electrode, given the activity of all the other electrodes. Let $\mathbf{X}_{-j}$, an (n-1) by T matrix, be the matrix of observations without the j^{th} electrode and $\mathbf{X}_j$ of dimension T, be the vector of the j^{th} electrode observations. With $\mathbf{X}_{-j}$ and appropriately modified latent variable model mixture component parameters, we can infer the most probable mixture component at

each time point, $\omega_{-j,t}$. Let Ω_{-j} of dimension T, be the vector of the inferred mixture component indices. Using the joint normality of $\mathbf{X}_{-j}$ and $\mathbf{X}_j$ given Ω_{-j}, we can use $E[\mathbf{X}_j|\mathbf{X}_{-j},\Omega_{-j}]$ as our prediction of the j^{th} electrode observations. The prediction error is defined as the sum-of-squared differences between the model prediction and the observed activity across electrodes and time points:

$$PE = \sum_{j=1}^{n} ||\mathbb{E}[\mathbf{X}_j|\mathbf{X}_{-j},\Omega_{-j}] - \mathbf{X}_j||. \tag{3.6}$$

This metric is similar to the metric introduced in Subsection 2.2.2. The difference is in the use here of Ω_{-j} instead of Ω in calculating prediction error; the "left-out" channel is used in inferring Ω.

3.2.5 Model Class Comparisons

Model comparisons for the model types discussed in Subsection 3.2.2 were made across subjects, bin widths (50, 100, and 200 milliseconds), percentage of training data used (10%, 50%, and 100%), number of model mixture components (2-4), and latent variable dimensionality (1-3) with respect to the metrics introduced in Subsection 3.2.4.

For comparing corresponding[6] individual models, the Wilcoxon paired-sample test was used. We tested for differences with the two-sided test, and in both directions with one-sided tests. In comparing the *NMI* for model pairs, test trials in each cross-validation fold were partitioned into as many sets as possible while ensuring that all 10 trial types were present in each set and the proportions of trial types was similar across the sets. *NMI* was then computed for each set. *NMI* across the sets and cross-validation folds was then compared. The *PE* was computed for each test trial in each cross-validation fold for model pairs and then the *PE* across trials and cross-validation

[6]For instance, a GMM_{diag} with 2 mixture components and a GMM_{full} with the same number of mixture components.

folds compared. Because there were many individual model comparisons, we controlled the false discovery rate with the Benjamini-Hochberg procedure [BH95], using a false discovery rate of 0.05.

There were two criteria for determining that a model class does better with respect to *PE* or *NMI*. The first criterion was that the percentage of significant (with a one-sided test and the two-sided test concurring) individual model comparison results for that model class is at least 5% more than the percentage of significant results for the other.[7] The second criterion was that the result is consistent across a majority of the subjects.[8]

3.3 Results

We observed that GMM_{full} had greater *PE* than GMM_{diag}, MFA_{CIPN}, and MFA_{CVPN} when 10% of the available training data was used, but then had lower *PE* when 50% and 100% of the available training data was used. However, GMM_{diag}, MFA_{CIPN}, and MFA_{CVPN} had greater *NMI* regardless of the percentage of training data used. We observed a similar effect for GMM_{full}-MM and HMM_{full} (full covariance models) with the corresponding spatiotemporal models. Figure 3.3 is for NY441 with a time point bin width of 50 milliseconds, 3 mixture components, and where applicable, a latent variable dimensionality of 1; and is representative of the aforementioned result summarized in tables A.1 to A.9.

We also observed that the models without shared variance (diagonal covariance models) only had lower *NMI* than a model with shared variance—specifically constrained shared variance models with the CIPN constraint— when a Markov temporal constraint was incorporated into the model. Figure 3.4 is for NY453 with a time point bin width

[7] As a concrete example, if the GMM_{diag} class does better than GMM_{full} in 35% of the model comparisons with respect the *NMI*, while GMM_{full} does better in 30% of model comparisons, then our determination is that the GMM_{diag} class as a whole does better.

[8] Results that did not satisfy the second criterion were omitted from the tables in appendix A.

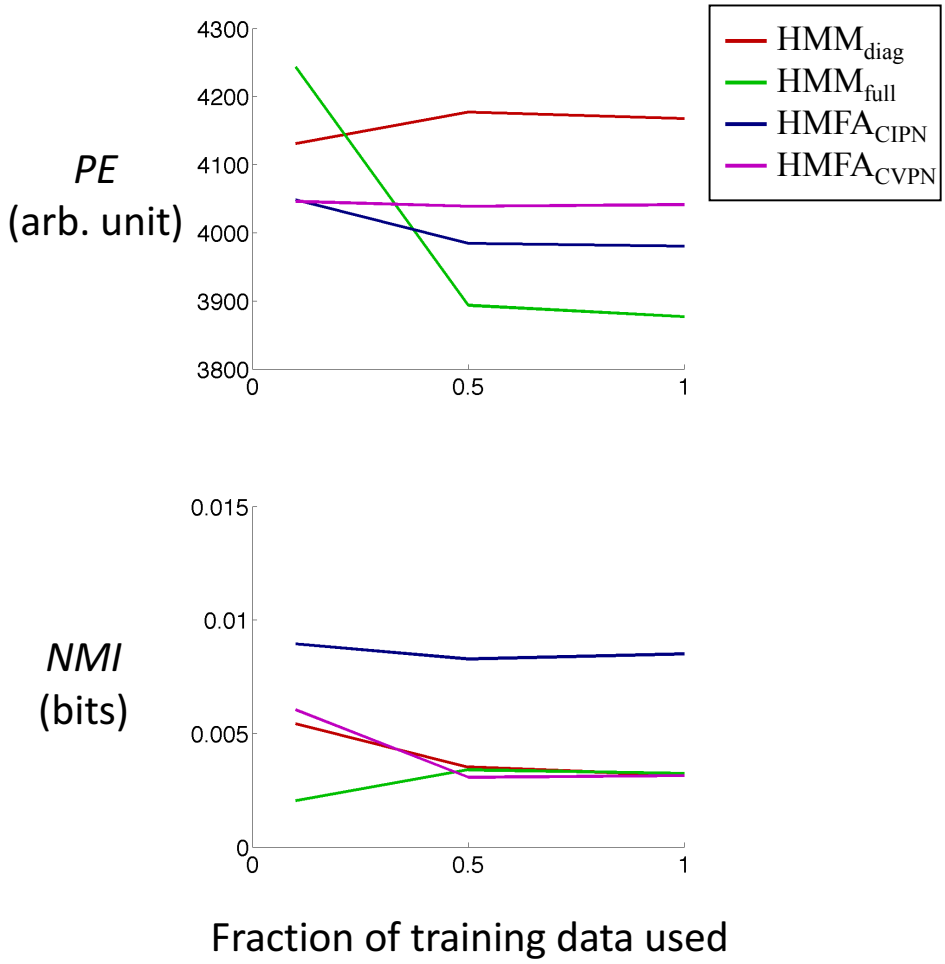


Figure 3.3: Prediction error and normalized mutual information of HMM_{diag}, HMM_{full}, $HMFA_{CIPN}$, and $HMFA_{CVPN}$ with 10%, 50%, and 100% of training data used for subject NY441 with a time point bin width of 50 milliseconds, 3 mixture components, and where applicable, a latent variable dimensionality of 1. The *PE* shown here is the median of medians of the *PE*s for the trials in the cross-validation folds. The *NMI* shown here is the median of the *NMI*s for the cross-validation folds.

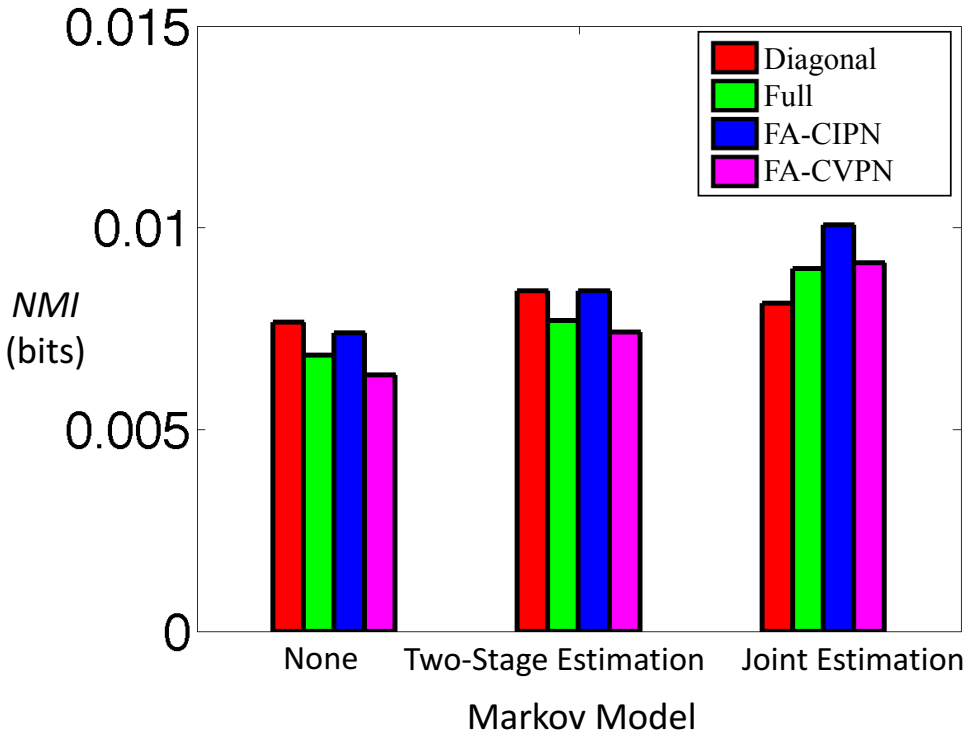


Figure 3.4: Normalized mutual information of all model types with 100% of training data used for subject NY453. The *NMI* shown here is the median of the of the *NMI*s for the cross-validation folds.

of 50 milliseconds, 3 mixture components, and where applicable, a latent variable dimensionality of 1; and is representative of the aforementioned result summarized in tables A.12 and A.13.

Another finding was that adding the Markov temporal constraint led to an increase in *NMI* with the joint estimation resulting in higher *NMI* than the two-stage estimation. Figure 3.4 is representative of this result which is summarized in Table A.14. Finally, $\text{HMFA}_{\text{CIPN}}$ had the greatest *NMI* among all the models. This result is detailed in tables A.13 and A.14.

3.4 Discussion

Of the two metrics, *NMI* and *PE*, employed in our analyses, only *NMI* leverages the stimulus-behavior information of the dataset; *PE* is more directly connected with the covariance. We found that while full covariance models are better able than the other models to capture covariance in the Cross-Modal Speech neural activity with increasing training data, much of the additional covariance captured does not seem to be connected with the stimulus-behavior information.

Even with the constrained shared variance models, there was an advantage relative to the diagonal covariance methods in modeling stimulus-behavior information only with the addition of a Markov temporal constraint. This advantage was specific to the constrained shared variance model with the assumption of component-invariant private noise. The question of whether this is due to a fundamental property of the ECoG signal or simply the result of a certain degree of limited data availability arises. A more thorough investigation would require access to more data as well as ensuring that the model complexity in terms of total parameter element count is matched for the CIPN and CVPN model variants.

Introducing the Markov temporal constraint resulted in improved modeling of stimulus-behavior information for all models, with the joint estimation further improving performance. Interestingly, the CIPN model variant only does better than CVPN variant with respect to modeling both the stimulus-behavior information and covariance with joint estimation. (Please see tables A.10, A.12 and A.13.)

Acknowledgement

Chapter 3, in part, is currently being prepared for submission for publication of the material. Omigbodun, A., Doyle, W. K., Devinsky, O., Friedman, D., Melloni, L., Thesen, T., Gilja, V. The dissertation author was the primary investigator and author of this material.

Chapter 4

Clustering multivariate electrocorticographic signal derived sequences with mixture of hidden Markov factor analyzer models

We introduce a Hidden-Markov Factor Analysis (HMFA)-based approach for clustering multivariate data sequences. We compare this novel HMFA-based approach with analogous hidden Markov model (HMM)-based methods. Experimental results from electrocorticographic (ECoG) signal trial clustering suggest that the appropriately constrained HMFA-based technique has advantages over the HMM-based methods in the context of data limited multivariate sequences.

4.1 Introduction

There is a growing need for techniques that address the problem of modeling sequential multivariate data, particularly in scenarios where the dataset size is limited. Indeed, such tools would be of utility in the neurophysiology and neuroengineering applications touched on in Chapter 1. In the case where access to brain signals is achieved through clinically invasive procedures and through patient interactions, there may be significant constraints on the quantity of data that can be obtained.

One approach to analyzing multivariate sequence data is to evaluate potentially salient dynamics by applying an unsupervised clustering based approach. The problem of clustering sequences with mixtures of hidden Markov models (HMMs) [Smy97] has been considered before. However, the application of this class of models with limited data availability introduces a significant challenge. Thus, we develop and evaluate methods for clustering multivariate sequences with a Hidden-Markov Factor Analysis [MM12] based approach in a data-limited context. Factor analysis is well-established as a statistical tool for modeling covariance structure with fewer degrees of freedom than assuming a full covariance matrix [GH96]. (Please see Equation 3.3 and the enveloping discussion.)

In our presentation here, we first introduce a Mixture of Hidden Markov Factor Analyzers (MHMFA) and we consider two variants of MHMFA. These variants are extensions of the HMFA models introduced in Subsection 3.2.2. Then, we touch on a general approach—based on the methods in [Smy97]—to initializing the mixture models considered here. We discuss the experiments that we use in assessing the different mixture models. Lastly, we present our results, which suggest that the more successful MHMFA variant is the extension of the HMFA model that performs best in Chapter 3.

4.2 Methods

4.2.1 Mixture of Hidden Markov Factor Analyzers (MHMFA)

We describe the MHMFA latent variable model as follows. Let $\mathbf{x}_{it} \in \mathbb{R}^r$ represent the multivariate observation at time step $t \in \{1,...,T_i\}$ of data sequence $\mathbf{x}_i$, $i \in 1,...,n$. Let $\mathbf{z}_{it} \in \mathbb{R}^s$ ($s < r$), and $\omega_{it} \in \{1,...,m\}$ respectively denote the factor and Markov state corresponding to $\mathbf{x}_{it}$. Just as in the formulation of the mixture of factor analyzers [GH96], we have the following linear-Gaussian relationship for each factor analyzer:

$$P(\mathbf{x}_{it}|\mathbf{z}_{it}, \omega_{it} = j) = \mathcal{N}(\Lambda_j \mathbf{z}_{it} + \mu_j, \Sigma_j), \tag{4.1}$$

where $\Lambda_j \in \mathbb{R}^{r \times s}$, $\mu_j \in \mathbb{R}^r$, and the covariance matrix, $\Sigma_j \in \mathbb{R}^{r \times r}$ ($j \in \{1,...,m\}$) are model parameters that are learned. As in standard Factor Analysis, Σ_j is constrained to be diagonal (with the diagonal entries representing the private noise or variance in each observed dimension) and

$$P(\mathbf{z}_{it}) = \mathcal{N}(0, I). \tag{4.2}$$

The factor analyzers at different time points are linked through an m-state HMM with start probability vector, π, and transition probability matrix, A, parameters that are also learned:

$$\pi_j = P(\omega_{i1} = j), \;\; A_{jk} = P(\omega_{it} = k|\omega_{i,t-1} = j), \;\; j,k \in \{1,...,m\}. \tag{4.3}$$

This completes the description of the HMFA (subsection 2.2.1) model:

$$\mathcal{M} = \{\pi, A, (\mu_j, \Lambda_j, \Sigma_j)_{j=1}^m\}. \tag{4.4}$$

Now we can consider the case where we have a finite set of c HMFA models with the h-th model:

$$\mathcal{M}^h = \{\pi^h, A^h, (\mu_j^h, \Lambda_j^h, \Sigma_j^h)_{j=1}^m\}, \;\; h \in \{1, ..., c\}, \tag{4.5}$$

and a probability vector ρ that parametrizes the mixing proportions:

$$\rho_h = P(\nu_h = 1). \tag{4.6}$$

ν_h is an indicator variable that equals 1 when a sequence $\mathbf{x}_i$ is generated by $\mathcal{M}^h$ and 0 otherwise. The complete MHMFA model is as follows:

$$\mathcal{M} = \{\rho, (\pi^h, A^h, (\mu_j^h, \Lambda_j^h, \Sigma_j^h)_{j=1}^m)_{h=1}^c\}. \tag{4.7}$$

Just like HMFA, MHMFA has its own AECM algorithm for parameter estimation and we outline the algorithm below.

First E-Step In this step, the factors $\mathbf{z}_{it}$ are marginalized out and only the state labels remain as missing data in the expected log-likelihood:

$$\begin{aligned} l_1(\mathcal{M}) = \sum_{i=1}^{n} \sum_{h=1}^{c} \mathbb{E}[\nu_h|\mathbf{x}_i] \Bigg(& \log \rho_h + \sum_{j=1}^{m} P(\omega_{i1}^h = j|\mathbf{x}_i) \log \pi_j^h \\ & + \sum_{t=2}^{T_i} \sum_{j=1}^{m} \sum_{k=1}^{m} P(\omega_{it}^h = k|\omega_{i,t-1}^h = j, \mathbf{x}_i) \log A_{jk}^h \\ & + \sum_{t=1}^{T_i} \sum_{j=1}^{m} P(\omega_{it}^h = j|\mathbf{x}_i) \log P_h(\mathbf{x}_{it}|\omega_{it}^h = j) \Bigg), \end{aligned} \tag{4.8}$$

where

$$\mathbb{E}[\nu_h|\mathbf{x}_i] \propto P(\mathbf{x}_i, \nu_h = 1) = \rho_h P(\mathbf{x}_i|\nu_h = 1). \tag{4.9}$$

$P(\omega_{it}^h = j|\mathbf{x}_i)$ and $P(\omega_{it}^h = k|\omega_{i,t-1}^h = j,\mathbf{x}_i)$ are calculated using the forward-backward algorithm for HMMs, and $P_h(\mathbf{x}_{it}|\omega_{it}^h = j)$ can be determined from Equations 4.1 and 4.2:

$$\begin{aligned}&\log P_h(\mathbf{x}_{it}|\omega_{it}^h = j) \propto \\ &-\frac{1}{2}\log|\Lambda_j^h(\Lambda_j^h)' + \Sigma_j^h| - \frac{1}{2}(\mathbf{x}_{it} - \mu_j^h)'(\Lambda_j^h(\Lambda_j^h)' + \Sigma_j^h)^{-1}(\mathbf{x}_{it} - \mu_j^h).\end{aligned} \tag{4.10}$$

First CM-Step $l_1(\mathcal{M})$ is maximized with respect to ρ and $(\pi^h, A^h, (\mu_j^h)_{j=1}^m)_{h=1}^c$ to obtain the following update rules:

$$\begin{aligned}\rho_h &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}[\nu_h|\mathbf{x}_i] \\ \mu_j^h &= \frac{\sum_{i=1}^n \mathbb{E}[\nu_h|\mathbf{x}_i]\sum_{t=1}^{T_i} P(\omega_{it}^h = j|\mathbf{x}_i)\mathbf{x}_{it}}{\sum_{i=1}^n \mathbb{E}[\nu_h|\mathbf{x}_i]\sum_{t=1}^{T_i} P(\omega_{it}^h = j|\mathbf{x}_i))} \\ \pi_j^h &= \frac{\sum_{i=1}^n \mathbb{E}[\nu_h|\mathbf{x}_i]P_h(\omega_{i1}^h = j|\mathbf{x}_i)}{n} \\ A_{jk}^h &= \frac{\sum_{i=1}^n \mathbb{E}[\nu_h|\mathbf{x}_i]\sum_{t=2}^{T_i} P(\omega_{it}^h = k|\omega_{i,t-1}^h = j,\mathbf{x}_i)}{\sum_{i=1}^n \mathbb{E}[\nu_h|\mathbf{x}_i]\sum_{t=2}^{T_i}\sum_{k=1}^m P(\omega_{it}^h = k|\omega_{i,t-1}^h = j,\mathbf{x}_i)}\end{aligned} \tag{4.11}$$

Second E-Step In this step, both the factors and state labels are taken as missing data in the expected log-likelihood:

$$\begin{aligned}l_2(\mathcal{M}) = \sum_{i=1}^n\sum_{h=1}^c \mathbb{E}[\nu_h|\mathbf{x}_i]\Bigg(&\log\rho_h + \sum_{j=1}^m P(\omega_{i1}^h = j|\mathbf{x}_i)\log\pi_j^h \\ &+ \sum_{t=2}^{T_i}\sum_{j=1}^m\sum_{k=1}^m P(\omega_{it}^h = k|\omega_{i,t-1}^h = j,\mathbf{x}_i)\log A_{jk}^h \\ &+ \sum_{t=1}^{T_i}\sum_{j=1}^m P(\omega_{it}^h = j|\mathbf{x}_i)\mathbb{E}[\log P_h(\mathbf{x}_{it}|\mathbf{z}_{it},\omega_{it}^h = j)|\mathbf{x}_i] \\ &+ \sum_{t=1}^{T_i}\sum_{j=1}^m P(\omega_{it}^h = j|\mathbf{x}_i)\log P(\mathbf{z}_{it})\Bigg),\end{aligned} \tag{4.12}$$

where:

$$\begin{aligned}&\mathbb{E}[\log P_h(\mathbf{x}_{it}|\mathbf{z}_{it},\omega_{it}^h=j)|\mathbf{x}_i]\propto\\&-\frac{1}{2}\log|\Sigma_j^h|-\frac{1}{2}\mathbb{E}[(\mathbf{x}_{it}-\mu_j^h-\Lambda_j^h\mathbf{z}_{it})'(\Sigma_j^h)^{-1}(\mathbf{x}_{it}-\mu_j^h-\Lambda_j^h\mathbf{z}_{it})|\mathbf{x}_i].\end{aligned}\tag{4.13}$$

Second CM-Step With the estimates from the first CM-Step we have the following updates for $((\Lambda_j^h,\Sigma_j^h)_{j=1}^m)_{h=1}^c$:

$$\begin{aligned}\Lambda_j^h&=\sum_{i=1}^n\mathbb{E}[\nu_h|\mathbf{x}_i]\sum_{t=1}^{T_i}P(\omega_{it}^h=j|\mathbf{x}_i)\cdot(\mathbf{x}_{it}-\mu_j^h)\cdot\mathbb{E}_h[\mathbf{z}_{it}{}'|\mathbf{x}_{it},\omega_{it}^h=j]\\&\quad\cdot\left(\sum_{i=1}^n\mathbb{E}[\nu_h|\mathbf{x}_i]\sum_{t=1}^{T_i}P(\omega_{it}^h=j|\mathbf{x}_i)\cdot\mathbb{E}_h[\mathbf{z}_{it}\mathbf{z}_{it}{}'|\mathbf{x}_{it},\omega_{it}^h=j]\right)^{-1}\\\Sigma_j^h&=\frac{1}{\sum_{i=1}^n\mathbb{E}[\nu_h|\mathbf{x}_i]\sum_{t=1}^{T_i}P(\omega_{it}^h=j|\mathbf{x}_i)}\\&\quad\cdot\operatorname{diag}\left[\sum_{i=1}^n\mathbb{E}[\nu_h|\mathbf{x}_i]\sum_{t=1}^{T_i}P(\omega_{it}^h=j|\mathbf{x}_i)\left((\mathbf{x}_{it}-\mu_j^h)(\mathbf{x}_{it}-\mu_j^h)'\right.\right.\\&\qquad\left.\left.-\Lambda_j^h\mathbb{E}_h[\mathbf{z}_{it}|\mathbf{x}_{it},\omega_{it}^h=j](\mathbf{x}_{it}-\mu_j^h)'\right)\right],\end{aligned}\tag{4.14}$$

where:

$$\begin{aligned}&\mathbb{E}_h[\mathbf{z}_{it}|\mathbf{x}_{it},\omega_{it}^h=j]=\beta_j^h(\mathbf{x}_{it}-\mu_j^h)\\&\mathbb{E}_h[\mathbf{z}_{it}\mathbf{z}_{it}{}'|\mathbf{x}_{it},\omega_{it}=j]=I-\beta_j^h\Lambda_j^h+\beta_j^h(\mathbf{x}_{it}-\mu_j^h)(\mathbf{x}_{it}-\mu_j^h)'(\beta_j^h)'\\&\beta_j^h=(\Lambda_j^h)'\left(\Sigma_j^h+\Lambda_j^h(\Lambda_j^h)'\right)^{-1}.\end{aligned}\tag{4.15}$$

We will refer to the model, $\mathcal{M}$, in Equation 4.7 as MHMFA$_{\text{CVPN}}$ hereafter, with CVPN standing for Component-Variant Private Noise. Please note that all the mixture models introduced here are extensions of the Markov models treated in Subsection 3.2.2. We

also consider $\text{MHMFA}_{\text{CIPN}}$:

$$\mathcal{M} = \{\rho, (\pi^h, A^h, (\mu_j^h, \Lambda_j^h)_{j=1}^m, \Sigma^h)_{h=1}^c\}, \tag{4.16}$$

in which the private noise parameter Σ^h is invariant among the factor analyzers within each underlying HMFA model. The intuition behind considering a model in which private noise is invariant across the hidden discrete states is that while the covariance structure may shift over time, for some problems we may expect the private noise characteristics to remain fixed. For example, if private noise is interpreted as sensor specific data acquisition noise, it is reasonable to expect the private noise variance to be constant across time. For this variant, the update rule for Σ^h is as follows:

$$\begin{aligned}\Sigma^h = & \frac{1}{\sum_{h=1}^c \sum_{i=1}^n \mathbb{E}[\nu_h|\mathbf{x}_i] \sum_{t=1}^{T_i} P(\omega_{it}^h = j|\mathbf{x}_i)} \\ & \cdot \text{diag}\left[\sum_{h=1}^c \sum_{i=1}^n \mathbb{E}[\nu_h|\mathbf{x}_i] \sum_{t=1}^{T_i} P(\omega_{it}^h = j|\mathbf{x}_i)\left((\mathbf{x}_{it} - \mu_j^h)(\mathbf{x}_{it} - \mu_j^h)'\right.\right. \\ & \left.\left. - \Lambda_j^h \mathbb{E}_h[\mathbf{z}_{it}|\mathbf{x}_{it}, \omega_{it}^h = j](\mathbf{x}_{it} - \mu_j^h)'\right)\right].\end{aligned} \tag{4.17}$$

Furthermore, two HMM mixture model types—$\text{MHMM}_{\text{full}}$ and $\text{MHMM}_{\text{diag}}$—are considered in our analyses. The Gaussian covariance parameters are unconstrained for $\text{MHMM}_{\text{full}}$, but constrained to be diagonal for $\text{MHMM}_{\text{diag}}$.

4.2.2 Algorithm Initialization

Let $\mathcal{M}_i$ denote an underlying Markov model (HMM_{diag}, HMM_{full}, $\text{HMFA}_{\text{CIPN}}$, or $\text{HMFA}_{\text{CVPN}}$) fit to data sequence $\mathbf{x}_i, i \in 1, ..., n$. We constructed a symmetric log-

likelihood distance matrix, L, with each entry defined as:

$$L_{il} = \log P(\mathbf{x}_l|\mathcal{M}_i) + \log P(\mathbf{x}_i|\mathcal{M}_l) \tag{4.18}$$

as prescribed in [Smy97]. We then used L for (1) removing outlier sequences, and (2) selecting dissimilar underlying Markov models to initialize the components of the mixture model. We adopted these two steps described here (as the rest of the procedure laid out in [Smy97] did not produce satisfactory results).

Outlier Sequence Removal The intuition behind the algorithm here is to identify and eliminate—from model fitting—particular sequences that are dissimilar from a supermajority of the other sequences. In our work here, we used a threshold of $\lceil (2/3)n \rceil$. The procedure is as follows:

1. Compute the index corresponding to the minimum of each row of L.
2. Among those indices that correspond to row minima, determine those (not already identified) that occur more than a specified threshold number of rows.
3. If no indices cross the threshold, terminate the procedure. Otherwise, reduce the threshold number of rows by the number of newly identified indices that cross the threshold and repeat steps 1-3.

Dissimilar Markov Models Selection The goal of Algorithm 1 is to identify a subset of models that are the most dissimilar from one another.

Algorithm 1 Dissimilar Markov Models Selection

Input: log-likelihood matrix L, number of sequences n, number of mixture components m
$[i_1, i_2] = \arg\min_{1 \leq i < l \leq n} L_{il}$
$S = \{i_1, i_2\}$
for $j = 3$ **to** m **do**
  $i_j = \arg\min_{1 \leq i \leq n, i \notin S} \sum_{l \in S} L_{il}$
  $S = S \cup \{i_j\}$
end for
return S

4.2.3 Cross-modal Speech Experiment and ECoG Recordings

The relevant Cross-modal Speech Experiment and ECoG signal processing information are contained in Subsection 3.2.1.[1]

4.2.4 Model Experiments

For our work here, we performed a one-way Analysis of Variance for each electrode across samples from each trial type, and used the p-values to rank electrodes. We performed analyses with the 5 top-ranked electrodes (having the 5 lowest p-values) as well as the 10 top-ranked electrodes for each subject. Even after limiting the number of electrodes to 5 and 10, we experienced significant model fitting issues[2] with using the HMM_{full}-based initialization for $\text{MHMM}_{\text{full}}$. This gives a sense of how data-limited we were; we used 100% of the available training data in all analyses. We used the $\text{HMFA}_{\text{CIPN}}$-based initialization instead for $\text{MHMM}_{\text{full}}$ as it performed better than the analagous $\text{HMFA}_{\text{CVPN}}$-based initialization. The initial $\text{MHMM}_{\text{full}}$ Markov process covariance matrices, $\Sigma^h_{j,\text{full}}$, are constructed as follows:

$$\Sigma^h_{j,\text{full}} = \Lambda^h_j (\Lambda^h_j)' + \Sigma^h, \;\; j \in \{1, ..., m\}, \;\; h \in \{1, ..., c\} \tag{4.19}$$

[1] We used only 50-millisecond neural signal time steps in our analyses here.
[2] These model fitting issues were due to the occurrence of singular matrices.

In all pertinent analyses herein, the dimensionality, s, of the factors, $\mathbf{z}_{it}$, was set to 1. We considered underlying Markov models with 2 and 3 hidden Markov states. We conducted two experiments.

First Experiment We compared $\text{MHMM}_{\text{diag}}$, $\text{MHMM}_{\text{full}}$, $\text{MHMFA}_{\text{CIPN}}$, and $\text{MHMFA}_{\text{CVPN}}$ in a two-cluster problem on Cross-Modal Speech Experiment trials with and without audio. Trials with phase-scrambled audio were excluded from this analysis. We evaluated the models in terms of accuracy and normalized mutual information, *NMI*, defined as:

$$NMI = \frac{I(U;V)}{H(V)} \tag{4.20}$$

where U corresponds to trial cluster assignments and V corresponds to trial labels (audio present or absent). Since this was a two-cluster problem, the number of component Markov models in the mixture models was set to 2.

Second Experiment All ten Cross-Modal Speech Experiment trial types were considered in this comparison of $\text{MHMM}_{\text{diag}}$, $\text{MHMM}_{\text{full}}$, $\text{MHMFA}_{\text{CIPN}}$, and $\text{MHMFA}_{\text{CVPN}}$. We evaluated the models in terms of normalized mutual information, *NMI*, with U corresponding to trial cluster assignments and V to an index between 1 and 10. We considered Markov mixture models with 2, 3, and 4 component Markov models.

4.2.5 Model Fitting and Class Comparisons

We employed 4-fold cross-validation in model fitting. Model comparisons for the model types discussed in Subsection 4.2.1 were made across subjects, the number of electrodes, the number of Markov process hidden states, and the number of component Markov models in the mixture model, with respect to accuracy and *NMI*. In statistical tests comparing the accuracy and *NMI* of models, test trials in each cross-validation fold

were partitioned into as many sets as possible while ensuring that all trials types were present in each set, and the proportions of trial types were similar across the sets. The accuracy and *NMI* were then computed for each set, and then the values across the sets and cross-validation folds were compared in a Wilcoxon paired-sample test. We only consider the comparisons between $\text{MHMM}_{\text{full}}$ and $\text{MHMFA}_{\text{CIPN}}$ (and not $\text{MHMM}_{\text{full}}$ with the other model classes) because $\text{MHMM}_{\text{full}}$ and $\text{MHMFA}_{\text{CIPN}}$ share the same initialization stage in our work here. The criteria for determining whether a model class performed better with respect to accuracy and/or *NMI* were the same as in Subsection 3.2.5.

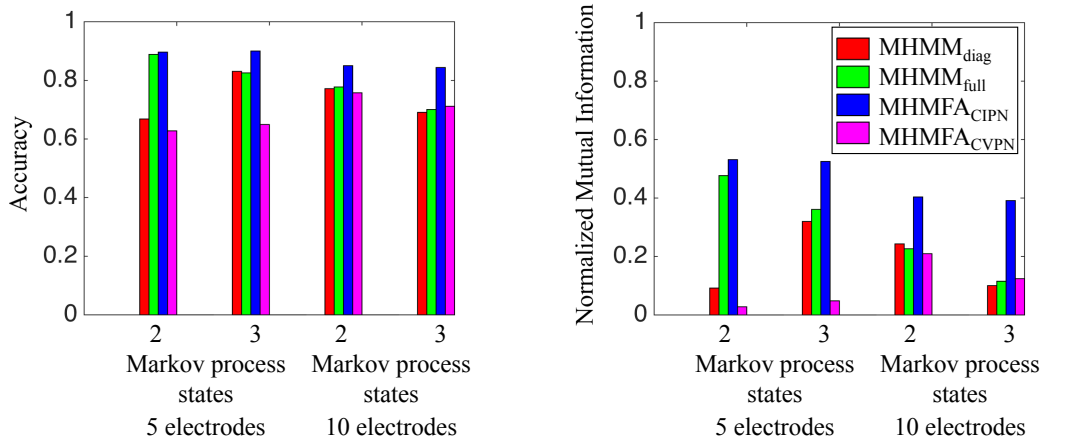
4.3 Results

For the first experiment, model comparisons were made across patients, the number of electrodes, and the number of Markov process hidden states (Wilcoxon paired-sample test, Benjamini-Hochberg false discovery rate of 0.05). The results in Table 4.1 are for accuracy and normalized mutual information. Figure 4.1 shows the accuracy and normalized mutual information results for subject NY455 across the number of electrodes, and the number of Markov process hidden states. This figure is representative of the results summarized in Table 4.1.

In the second experiment, we evaluated models in terms of normalized mutual information, and made comparisons across patients, the number of electrodes, the number of Markov process hidden states, and the number of component Markov models in the mixture model. The results are shown in Table 4.2. Figure 4.2 shows the normalized mutual information results for NY441 across the number of electrodes, the number of Markov process hidden states, and the number of Markov models in the mixture. This figure is representative of the results summarized in Table 4.2.

Table 4.1: Model class comparisons

Accuracy				
Model Class			Outcome	
A	B	Total number of model comparisons	A statistically significantly greater	B statistically significantly greater
$\text{MHMFA}_{\text{CIPN}}$	$\text{MHMM}_{\text{full}}$	24	8 (33.33%)	0 (0%)
$\text{MHMFA}_{\text{CIPN}}$	$\text{MHMFA}_{\text{CVPN}}$	23	5 (21.74%)	0 (0%)
$\text{MHMFA}_{\text{CIPN}}$	$\text{MHMM}_{\text{diag}}$	24	4 (16.67%)	0 (0%)
$\text{MHMM}_{\text{diag}}$	$\text{MHMFA}_{\text{CVPN}}$	23	0 (0%)	0 (0%)
Normalized Mutual Information				
Model Class			Outcome	
A	B	Total number of model comparisons	A statistically significantly greater	B statistically significantly greater
$\text{MHMFA}_{\text{CIPN}}$	$\text{MHMM}_{\text{full}}$	24	8 (33.33%)	0 (0%)
$\text{MHMFA}_{\text{CIPN}}$	$\text{MHMFA}_{\text{CVPN}}$	23	8 (34.78%)	0 (0%)
$\text{MHMFA}_{\text{CIPN}}$	$\text{MHMM}_{\text{diag}}$	24	4 (16.67%)	0 (0%)
$\text{MHMM}_{\text{diag}}$	$\text{MHMFA}_{\text{CVPN}}$	23	0 (0%)	0 (0%)



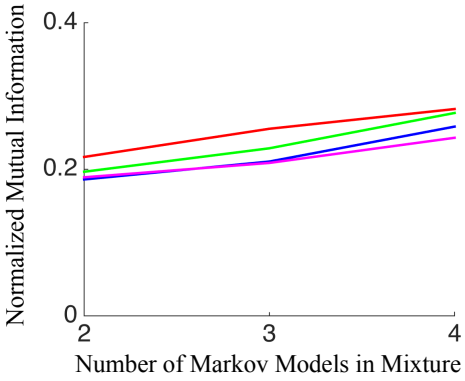
(a) Median accuracy

(b) Median normalized mutual information

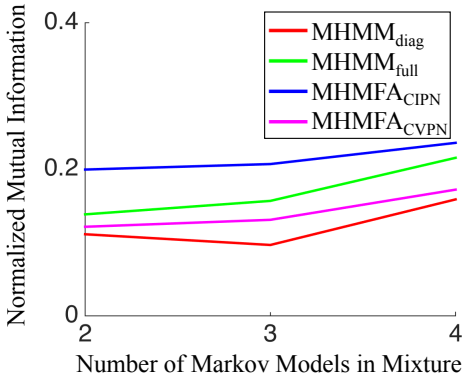
Figure 4.1: Results of the first experiment for subject NY455

Table 4.2: Model class normalized mutual information comparisons

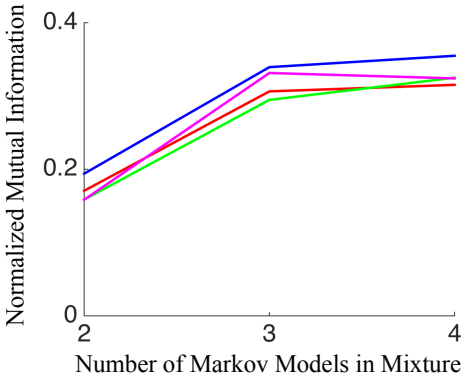
Model Class			Outcome	
A	B	Total number of model comparisons	A statistically significantly greater	B statistically significantly greater
$\text{MHMFA}_{\text{CIPN}}$	$\text{MHMM}_{\text{full}}$	72	13 (18.06%)	0 (0%)
$\text{MHMFA}_{\text{CIPN}}$	$\text{MHMFA}_{\text{CVPN}}$	71	28 (39.44%)	0 (0%)
$\text{MHMFA}_{\text{CIPN}}$	$\text{MHMM}_{\text{diag}}$	72	15 (20.83%)	0 (0%)
$\text{MHMM}_{\text{diag}}$	$\text{MHMFA}_{\text{CVPN}}$	71	1 (1.41%)	0 (0%)



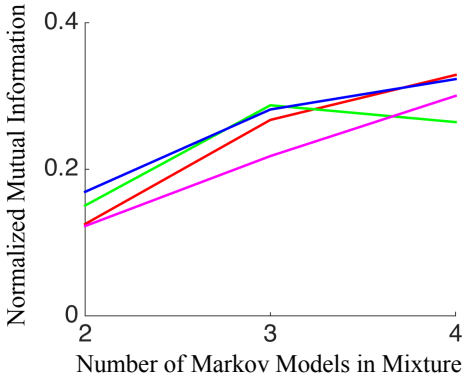
(a) 5 electrodes, 2 Markov process states



(b) 5 electrodes, 3 Markov process states



(c) 10 electrodes, 2 Markov process states



(d) 10 electrodes, 3 Markov process states

Figure 4.2: Results of the second experiment for NY441

4.4 Discussion

In the first experiment, we found that $\text{MHMFA}_{\text{CIPN}}$ outperformed the other model classes with respect to accuracy and normalized mutual information. However, we did not observe a difference between $\text{MHMFA}_{\text{CVPN}}$ and $\text{MHMM}_{\text{diag}}$. This suggests that while there may be benefits to modeling changes over time in shared variance between electrode channels, the additional model complexity of a private noise parameter for each Markov process Gaussian component disadvantaged $\text{MHMFA}_{\text{CVPN}}$ relative to $\text{MHMFA}_{\text{CIPN}}$, which has a single private noise parameter shared by the Gaussian components within each Markov model. This finding is consistent not only with our observations in Section 3.4 but also with our intuition about the nature of ECoG recording[3] that the electrodes' shared variance has a more reliable connection than private noise with that underlying neural modulation. As we mentioned in Section 3.4, more work is necessary to make any determination in this matter.

We also note that $\text{MHMFA}_{\text{CIPN}}$ did better than $\text{MHMM}_{\text{full}}$ even though the models had the same initialization stage in our work here, and that there were model fitting issues albeit very limited with $\text{MHMFA}_{\text{CVPN}}$[4]. These seem to point to how data-limited we are. Finally, the main results of the second experiment, in which the number of trial conditions was expanded from two to ten, match those of the first experiment very closely.

Acknowledgement

Chapter 4, in part, has been submitted for publication of the material as it may appear in Advances in Neural Information Processing Systems, 2017, Omigbodun, A.,

[3] Each individual recording electrode is subject to substantial external electromagnetic interference uncorrelated to underlying neural modulation.

[4] Note that there are slight discrepancies in the numbers of model comparisons detailed in Tables 4.1 and 4.2 because of these fitting issues due to the occurrence of singular matrices.

Doyle, W. K., Devinsky, O., Friedman, D., Melloni, L., Thesen, T., Gilja, V. The dissertation author was the primary investigator and author of this paper.

Chapter 5

Conclusion and Future Directions

Our work on trial clustering in Chapter 4 serves as a proof of concept of an algorithmic BMI system component for identifying stimuli and behavior with salient ECoG neural responses. The current paradigm in studies with neuroprosthetic assistive devices is to identify suitable neural responses for BMI system control during a period of instructed behavior. But what if that period of instructed behavior is replaced with period of free behavior during which a BMI system component identifies behavior that has a consistent and salient neural activation pattern?

Additionally, the model framework laid out in chapters 3 and 4 could serve as a useful tool for understanding the connections between aspects of perception and cognition, and different components of ECoG neural signal variability. For instance, does the result about the private variance constraint tell us something about the nature of the ECoG signal or is it just a reflection of how data-limited we are? If the constraint does tell us something fundamental, what is the scope in terms of different aspects of perception, behavior, and cognition?

Thirdly, the advancement of BMI technology and neurophysiology that the methods proposed in this dissertation may enable could lead to new clinical diagnostic and

therapeutic techniques.

Appendix A

Model Class Comparison Tables

The results of the analyses in 3 are detailed in the following tables.

Table A.1: Model class comparison of GMM_{diag} and GMM_{full}

Prediction Error			
Percentage of training data used	Total number of model comparisons	GMM_{diag} statistically significantly smaller	GMM_{full} statistically significantly smaller
10%	54	40 (74.07%)	11 (20.37%)
50%	54	1 (1.85%)	49 (90.74%)
100%	54	0 (0%)	53 (98.15%)
Normalized Mutual Information			
Percentage of training data used	Total number of model comparisons	GMM_{diag} statistically significantly greater	GMM_{full} statistically significantly greater
10%	54	45 (83.33%)	0 (0%)
50%	54	30 (55.56%)	0 (0%)
100%	54	23 (42.59%)	0 (0%)

Table A.2: Model class comparison of MFA_{CVPN} and GMM_{full}

Prediction Error				
Percentage of training data used	Latent variable dimensionality of MFA_{CVPN}	Total number of model comparisons	MFA_{CVPN} statistically significantly smaller	GMM_{full} statistically significantly smaller
10%	1	54	47 (87.04%)	3 (5.56%)
50%	1	54	9 (16.67%)	41 (75.93%)
100%	1	54	1 (1.85%)	49 (90.74%)
10%	2	54	51 (94.44%)	0 (0%)
50%	2	54	12 (22.22%)	37 (68.52%)
100%	2	54	4 (7.41%)	45 (83.33%)
10%	3	54	52 (96.30%)	0 (0%)
50%	3	54	15 (27.78%)	32 (59.26%)
100%	3	54	4 (7.41%)	43 (79.63%)
Normalized Mutual Information				
Percentage of training data used	Latent variable dimensionality of MFA_{CVPN}	Total number of model comparisons	MFA_{CVPN} statistically significantly greater	GMM_{full} statistically significantly greater
10%	1	54	47 (87.04%)	0 (0%)
50%	1	54	27 (50%)	0 (0%)
100%	1	54	27 (50%)	0 (0%)
10%	2	54	46 (85.19%)	0 (0%)
50%	2	54	25 (46.30%)	0 (0%)
100%	2	54	23 (42.59%)	0 (0%)
10%	3	54	47 (87.04%)	0 (0%)
50%	3	54	27 (50%)	0 (0%)
100%	3	54	23 (42.59%)	0 (0%)

Table A.3: Model class comparison of MFA_{CIPN} and GMM_{full}

Prediction Error				
Percentage of training data used	Latent variable dimensionality of MFA_{CIPN}	Total number of model comparisons	MFA_{CIPN} statistically significantly smaller	GMM_{full} statistically significantly smaller
10%	1	54	51 (94.44%)	1 (1.85%)
50%	1	54	9 (16.67%)	41 (77.78%)
100%	1	54	2 (3.70%)	50 (92.59%)
10%	2	54	52 (96.30%)	0 (0%)
50%	2	54	13 (24.07%)	37 (68.52%)
100%	2	54	3 (5.56%)	46 (85.19%)
10%	3	54	52 (96.30%)	0 (0%)
50%	3	54	14 (25.93%)	34 (62.96%)
100%	3	54	3 (5.56%)	45 (83.33%)
Normalized Mutual Information				
Percentage of training data used	Latent variable dimensionality of MFA_{CIPN}	Total number of model comparisons	MFA_{CIPN} statistically significantly greater	GMM_{full} statistically significantly greater
10%	1	54	34 (62.96%)	3 (5.56%)
50%	1	54	27 (50%)	5 (9.26%)
100%	1	54	22 (40.74%)	7 (12.96%)
10%	2	54	36 (66.67%)	1 (1.85%)
50%	2	54	22 (40.74%)	4 (7.41%)
100%	2	54	24 (44.44%)	3 (5.56%)
10%	3	54	38 (70.37%)	1 (1.85%)
50%	3	54	25 (46.30%)	0 (0%)
100%	3	54	27 (50%)	2 (3.70%)

Table A.4: Model class comparison of GMM_{diag}-MM and GMM_{full}-MM

Prediction Error			
Percentage of training data used	Total number of model comparisons	GMM_{diag}-MM statistically significantly smaller	GMM_{full}-MM statistically significantly smaller
10%	54	40 (74.07%)	10 (18.52%)
50%	54	2 (3.70%)	50 (92.59%)
100%	54	0 (0%)	52 (96.30%)
Normalized Mutual Information			
Percentage of training data used	Total number of model comparisons	GMM_{diag}-MM statistically significantly greater	GMM_{full}-MM statistically significantly greater
10%	54	20 (37.04%)	0 (0%)
50%	54	17 (31.48%)	0 (0%)
100%	54	15 (27.78%)	0 (0%)

Table A.5: Model class comparison of MFA_{CVPN}-MM and GMM_{full}-MM

Prediction Error				
Percentage of training data used	Latent variable dimensionality of MFA_{CVPN}-MM	Total number of model comparisons	MFA_{CVPN}-MM statistically significantly smaller	GMM_{full}-MM statistically significantly smaller
10%	1	54	48 (88.89%)	3 (5.56%)
50%	1	54	9 (16.67%)	42 (77.78%)
100%	1	54	3 (5.56%)	48 (88.89%)
10%	2	54	51 (94.44%)	0 (0%)
50%	2	54	12 (22.22%)	36 (66.67%)
100%	2	54	5 (9.26%)	44 (81.48%)
10%	3	54	52 (96.30%)	0 (0%)
50%	3	54	15 (27.78%)	32 (59.26%)
100%	3	54	5 (9.26%)	42 (77.78%)
Normalized Mutual Information				
Percentage of training data used	Latent variable dimensionality of MFA_{CVPN}-MM	Total number of model comparisons	MFA_{CVPN}-MM statistically significantly greater	GMM_{full}-MM statistically significantly greater
10%	1	54	24 (44.44%)	0 (0%)
50%	1	54	20 (37.04%)	0 (0%)
100%	1	54	15 (27.78%)	0 (0%)
10%	2	54	20 (37.04%)	0 (0%)
50%	2	54	17 (31.48%)	0 (0%)
100%	2	54	14 (25.93%)	0 (0%)
10%	3	54	22 (40.74%)	0 (0%)
50%	3	54	20 (37.04%)	0 (0%)
100%	3	54	13 (24.07%)	0 (0%)

Table A.6: Model class comparison of MFA_{CIPN}-MM and GMM_{full}-MM

Prediction Error				
Percentage of training data used	Latent variable dimensionality of MFA_{CIPN}-MM	Total number of model comparisons	MFA_{CIPN}-MM statistically significantly smaller	GMM_{full}-MM statistically significantly smaller
10%	1	54	48 (88.89%)	2 (3.70%)
50%	1	54	10 (18.52%)	40 (74.07%)
100%	1	54	4 (7.41%)	50 (92.59%)
10%	2	54	51 (94.44%)	0 (0%)
50%	2	54	14 (25.93%)	39 (72.22%)
100%	2	54	4 (7.41%)	45 (83.33%)
10%	3	54	52 (96.30%)	0 (0%)
50%	3	54	14 (25.93%)	34 (62.96%)
100%	3	54	4 (7.41%)	44 (81.48%)
Normalized Mutual Information				
Percentage of training data used	Latent variable dimensionality of MFA_{CIPN}-MM	Total number of model comparisons	MFA_{CIPN}-MM statistically significantly greater	GMM_{full}-MM statistically significantly greater
10%	1	54	23 (42.59%)	1 (1.85%)
50%	1	54	28 (51.85%)	2 (3.70%)
100%	1	54	26 (48.15%)	1 (1.85%)
10%	2	54	21 (38.89%)	0 (0%)
50%	2	54	25 (46.30%)	0 (0%)
100%	2	54	26 (48.15%)	0 (0%)
10%	3	54	20 (37.04%)	0 (0%)
50%	3	54	29 (53.70%)	0 (0%)
100%	3	54	25 (46.30%)	0 (0%)

Table A.7: Model class comparison of HMM_{diag} and HMM_{full}

Prediction Error			
Percentage of training data used	Total number of model comparisons	HMM_{diag} statistically significantly smaller	HMM_{full} statistically significantly smaller
10%	53	39 (73.58%)	12 (22.64%)
50%	54	1 (1.85%)	50 (94.44%)
100%	54	0 (0%)	54 (100%)
Normalized Mutual Information			
Percentage of training data used	Total number of model comparisons	HMM_{diag} statistically significantly greater	HMM_{full} statistically significantly greater
10%	53	21 (39.62%)	0 (0%)
50%	54	18 (33.33%)	0 (0%)
100%	54	8 (14.81%)	0 (0%)

Table A.8: Model class comparison of $HMFA_{CVPN}$ and HMM_{full}

Prediction Error				
Percentage of training data used	Latent variable dimensionality of $HMFA_{CVPN}$	Total number of model comparisons	$HMFA_{CVPN}$ statistically significantly smaller	HMM_{full} statistically significantly smaller
10%	1	53	47 (88.68%)	4 (7.55%)
50%	1	54	8 (14.81%)	44 (81.48%)
100%	1	54	1 (1.85%)	51 (94.44%)
10%	2	53	50 (94.34%)	2 (3.77%)
50%	2	54	13 (24.07%)	39 (72.22%)
100%	2	54	3 (5.56%)	49 (90.74%)
10%	3	53	51 (96.23%)	1 (1.89%)
50%	3	54	18 (33.33%)	28 (51.85%)
100%	3	54	4 (7.41%)	45 (83.33%)
Normalized Mutual Information				
Percentage of training data used	Latent variable dimensionality of $HMFA_{CVPN}$	Total number of model comparisons	$HMFA_{CVPN}$ statistically significantly greater	HMM_{full} statistically significantly greater
10%	1	53	23 (43.40%)	0 (0%)
50%	1	54	17 (31.48%)	0 (0%)
100%	1	54	7 (12.96%)	0 (0%)
10%	2	53	22 (41.51%)	0 (0%)
50%	2	54	19 (35.19%)	0 (0%)
100%	2	54	5 (9.26%)	0 (0%)
10%	3	53	18 (33.96%)	0 (0%)
50%	3	54	17 (31.48%)	0 (0%)
100%	3	54	5 (9.26%)	0 (0%)

Table A.9: Model class comparison of $HMFA_{CIPN}$ and HMM_{full}

Prediction Error				
Percentage of training data used	Latent variable dimensionality of $HMFA_{CIPN}$	Total number of model comparisons	$HMFA_{CIPN}$ statistically significantly smaller	HMM_{full} statistically significantly smaller
10%	1	53	46 (86.79%)	3 (5.66%)
50%	1	54	7 (12.96%)	42 (77.78%)
100%	1	54	1 (1.85%)	52 (96.30%)
10%	2	53	48 (90.57%)	3 (5.66%)
50%	2	54	13 (24.07%)	33 (61.11%)
100%	2	54	3 (5.56%)	47 (87.04%)
10%	3	53	50 (94.34%)	1 (1.89%)
50%	3	54	19 (35.19%)	31 (57.41%)
100%	3	54	3 (5.56%)	45 (83.33%)
Normalized Mutual Information				
Percentage of training data used	Latent variable dimensionality of $HMFA_{CIPN}$	Total number of model comparisons	$HMFA_{CIPN}$ statistically significantly greater	HMM_{full} statistically significantly greater
10%	1	53	21 (39.62%)	0 (0%)
50%	1	54	26 (48.15%)	0 (0%)
100%	1	54	20 (37.04%)	0 (0%)
10%	2	53	26 (49.06%)	0 (0%)
50%	2	54	32 (59.26%)	0 (0%)
100%	2	54	22 (40.74%)	0 (0%)
10%	3	53	27 (50.94%)	0 (0%)
50%	3	54	34 (62.96%)	0 (0%)
100%	3	54	26 (48.15%)	0 (0%)

Table A.10: Model class prediction error comparisons (spatial constraints)

Model Class			Outcome	
A	B	Latent variable dimensionality of factor analysis–based models, Total number of model comparisons	A statistically significantly smaller	B statistically significantly smaller
MFA_{CVPN}	GMM_{diag}	1, 162	162 (100%)	0 (0%)
MFA_{CVPN}	GMM_{diag}	2, 162	161 (99.38%)	0 (0%)
MFA_{CVPN}	GMM_{diag}	3, 162	162 (100%)	0 (0%)
MFA_{CIPN}	GMM_{diag}	1, 162	151 (93.21%)	0 (0%)
MFA_{CIPN}	GMM_{diag}	2, 162	152 (93.83%)	1 (0.62%)
MFA_{CIPN}	GMM_{diag}	3, 162	154 (95.06%)	1 (0.62%)
MFA_{CIPN}	MFA_{CVPN}	1-3, 486	231 (47.53%)	98 (20.16%)
MFA_{CVPN}-MM	GMM_{diag}-MM	1, 162	162 (100%)	0 (0%)
MFA_{CVPN}-MM	GMM_{diag}-MM	2, 162	162 (100%)	0 (0%)
MFA_{CVPN}-MM	GMM_{diag}-MM	3, 162	162 (100%)	0 (0%)
MFA_{CIPN}-MM	GMM_{diag}-MM	1, 162	148 (91.36%)	1 (0.62%)
MFA_{CIPN}-MM	GMM_{diag}-MM	2, 162	151 (93.21%)	2 (1.23%)
MFA_{CIPN}-MM	GMM_{diag}-MM	3, 162	152 (93.83%)	2 (1.23%)
$HMFA_{CVPN}$	HMM_{diag}	1, 162	162 (100%)	0 (0%)
$HMFA_{CVPN}$	HMM_{diag}	2, 162	162 (100%)	0 (0%)
$HMFA_{CVPN}$	HMM_{diag}	3, 162	162 (100%)	0 (0%)
$HMFA_{CIPN}$	HMM_{diag}	1, 162	149 (91.98%)	2 (1.23%)
$HMFA_{CIPN}$	HMM_{diag}	2, 162	151 (93.21%)	1 (0.62%)
$HMFA_{CIPN}$	HMM_{diag}	3, 162	156 (96.30%)	0 (0%)
$HMFA_{CIPN}$	$HMFA_{CVPN}$	1-3, 486	147 (30.25%)	0 (0%)

Table A.11: Model class prediction error comparisons (Markov temporal constraint)

Model Class			Outcome	
A	B	Latent variable dimensionality of factor analysis–based models, Total number of model comparisons	A statistically significantly smaller	B statistically significantly smaller
GMM_{full}-MM	GMM_{full}	none, 162	72 (44.44%)	4 (2.47%)
GMM_{diag}-MM	GMM_{diag}	none, 162	108 (66.67%)	1 (0.62%)
MFA_{CVPN}-MM	MFA_{CVPN}	1-3, 486	325 (66.87%)	0 (0%)
MFA_{CIPN}-MM	MFA_{CIPN}	1-3, 486	242 (49.79%)	15 (3.09%)
HMM_{full}	GMM_{full}	none, 161	71 (44.10%)	33 (20.50%)
HMM_{diag}	GMM_{diag}	none, 162	96 (59.26%)	1 (0.62%)
$HMFA_{CVPN}$	MFA_{CVPN}	1-3, 486	336 (69.14%)	19 (3.91%)
$HMFA_{CIPN}$	MFA_{CIPN}	1-3, 486	310 (63.79%)	43 (8.85%)
HMM_{full}	GMM_{full}-MM	none, 161	65 (40.37%)	33 (20.50%)
HMM_{diag}	GMM_{diag}-MM	none, 162	48 (29.63%)	2 (1.23%)
$HMFA_{CVPN}$	MFA_{CVPN}-MM	1-3, 486	275 (56.58%)	38 (7.82%)
$HMFA_{CIPN}$	MFA_{CIPN}-MM	1-3, 486	294 (60.49%)	40 (8.23%)

Table A.12: Model class normalized mutual information comparisons (spatial constraints) for models without the Markov temporal constraint

Model Class			Outcome	
A	B	Latent variable dimensionality of factor analysis–based models, Total number of model comparisons	A statistically significantly greater	B statistically significantly greater
GMM_{diag}	GMM_{full}	none, 162	98 (60.49%)	0 (0%)
MFA_{CVPN}	GMM_{full}	1, 162	101 (62.35%)	0 (0%)
MFA_{CVPN}	GMM_{full}	2, 162	94 (58.02%)	0 (0%)
MFA_{CVPN}	GMM_{full}	3, 162	97 (59.88%)	0 (0%)
MFA_{CIPN}	GMM_{full}	1, 162	83 (51.23%)	15 (9.26%)
MFA_{CIPN}	GMM_{full}	2, 162	82 (50.62%)	8 (4.94%)
MFA_{CIPN}	GMM_{full}	3, 162	90 (55.56%)	3 (1.85%)
MFA_{CVPN}	GMM_{diag}	1, 162	0 (0%)	2 (1.23%)
MFA_{CVPN}	GMM_{diag}	2, 162	0 (0%)	5 (3.09%)
MFA_{CVPN}	GMM_{diag}	3, 162	0 (0%)	7 (4.32%)
MFA_{CIPN}	GMM_{diag}	1, 162	3 (1.85%)	22 (13.58%)
MFA_{CIPN}	GMM_{diag}	2, 162	1 (0.62%)	11 (6.79%)
MFA_{CIPN}	GMM_{diag}	3, 162	0 (0%)	11 (6.79%)

Table A.13: Model class normalized mutual information comparisons (spatial constraints) for models with the Markov temporal constraint

Model Class			Outcome	
A	B	Latent variable dimensionality of factor analysis–based models, Total number of model comparisons	A statistically significantly greater	B statistically significantly greater
GMM_{diag}-MM	GMM_{full}-MM	none, 162	52 (32.10%)	0 (0%)
MFA_{CVPN}-MM	GMM_{full}-MM	1, 162	59 (36.42%)	0 (0%)
MFA_{CVPN}-MM	GMM_{full}-MM	2, 162	51 (31.48%)	0 (0%)
MFA_{CVPN}-MM	GMM_{full}-MM	3, 162	55 (33.95%)	0 (0%)
MFA_{CIPN}-MM	GMM_{full}-MM	1, 162	77 (47.53%)	4 (2.47%)
MFA_{CIPN}-MM	GMM_{full}-MM	2, 162	72 (44.44%)	0 (0%)
MFA_{CIPN}-MM	GMM_{full}-MM	3, 162	74 (45.68%)	0 (0%)
MFA_{CVPN}-MM	GMM_{diag}-MM	1, 162	0 (0%)	4 (2.47%)
MFA_{CVPN}-MM	GMM_{diag}-MM	2, 162	0 (0%)	4 (2.47%)
MFA_{CVPN}-MM	GMM_{diag}-MM	3, 162	0 (0%)	5 (3.09%)
MFA_{CIPN}-MM	GMM_{diag}-MM	1, 162	35 (21.60%)	8 (4.94%)
MFA_{CIPN}-MM	GMM_{diag}-MM	2, 162	30 (18.52%)	2 (1.23%)
MFA_{CIPN}-MM	GMM_{diag}-MM	3, 162	25 (15.43%)	1 (0.62%)
MFA_{CIPN}-MM	MFA_{CVPN}-MM	1-3, 486	112 (23.05%)	9 (1.85%)
HMM_{diag}	HMM_{full}	none, 161	47 (29.19%)	0 (0%)
$HMFA_{CVPN}$	HMM_{full}	1, 161	47 (29.19%)	0 (0%)
$HMFA_{CVPN}$	HMM_{full}	2, 161	46 (28.57%)	0 (0%)
$HMFA_{CVPN}$	HMM_{full}	3, 161	40 (24.84%)	0 (0%)
$HMFA_{CIPN}$	HMM_{full}	1, 161	67 (41.61%)	0 (0%)
$HMFA_{CIPN}$	HMM_{full}	2, 161	80 (49.69%)	0 (0%)
$HMFA_{CIPN}$	HMM_{full}	3, 161	87 (54.04%)	0 (0%)
$HMFA_{CVPN}$	HMM_{diag}	1, 162	7 (4.32%)	10 (6.17%)
$HMFA_{CVPN}$	HMM_{diag}	2, 162	4 (2.47%)	20 (12.35%)
$HMFA_{CVPN}$	HMM_{diag}	3, 162	0 (0%)	17 (10.49%)
$HMFA_{CIPN}$	HMM_{diag}	1, 162	49 (30.25%)	0 (0%)
$HMFA_{CIPN}$	HMM_{diag}	2, 162	54 (33.33%)	0 (0%)
$HMFA_{CIPN}$	HMM_{diag}	3, 162	46 (28.40%)	0 (0%)
$HMFA_{CIPN}$	$HMFA_{CVPN}$	1-3, 486	147 (30.25%)	0 (0%)

Table A.14: Model class normalized mutual information comparisons (Markov temporal constraint)

Model Class			Outcome	
A	B	Latent variable dimensionality of factor analysis–based models, Total number of model comparisons	A statistically significantly greater	B statistically significantly greater
GMM_{full}-MM	GMM_{full}	none, 162	114 (70.37%)	0 (0%)
GMM_{diag}-MM	GMM_{diag}	none, 162	129 (79.63%)	0 (0%)
MFA_{CVPN}-MM	MFA_{CVPN}	1-3, 486	418 (86.01%)	0 (0%)
MFA_{CIPN}-MM	MFA_{CIPN}	1-3, 486	411 (84.57%)	10 (6.17%)
HMM_{full}	GMM_{full}	none, 161	75 (46.58%)	0 (0%)
HMM_{diag}	GMM_{diag}	none, 162	85 (52.47%)	0 (0%)
$HMFA_{CVPN}$	MFA_{CVPN}	1-3, 486	388 (79.84%)	0 (0%)
$HMFA_{CIPN}$	MFA_{CIPN}	1-3, 486	411 (84.57%)	0 (0%)
HMM_{full}	GMM_{full}-MM	none, 161	20 (12.42%)	0 (0%)
HMM_{diag}	GMM_{diag}-MM	none, 162	27 (16.67%)	0 (0%)
$HMFA_{CVPN}$	MFA_{CVPN}-MM	1-3, 486	138 (28.40%)	1 (0.21%)
$HMFA_{CIPN}$	MFA_{CIPN}-MM	1-3, 486	212 (43.62%)	0 (0%)

Bibliography

[ASGA96] Amos Arieli, Alexander Sterkin, Amiram Grinvald, and AD Aertsen. Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. *Science*, 273(5283):1868, 1996.

[BH95] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, pages 289–300, 1995.

[BMJC13] Kristofer E. Bouchard, Nima Mesgarani, Keith Johnson, and Edward F. Chang. Functional organization of human sensorimotor cortex for speech articulation. *Nature*, 495(7441):327–332, Mar 2013.

[CKC+11] Anne K Churchland, Roozbeh Kiani, Rishidev Chaudhuri, Xiao-Jing Wang, Alexandre Pouget, and Michael N Shadlen. Variance as a signature of neural computations during decision making. *Neuron*, 69(4):818–831, 2011.

[CY14] John P. Cunningham and Byron M. Yu. Dimensionality reduction for large-scale neural recordings. *Nat Neurosci*, 17(11):1500–1509, Nov 2014. Review.

[GH96] Zoubin Ghahramani and Geoffrey E Hinton. The em algorithm for mixtures of factor analyzers. Technical report, Technical Report CRG-TR-96-1, University of Toronto, 1996.

[JK09] Joshua Jacobs and Michael J Kahana. Neural representations of individual stimuli in humans revealed by gamma-band electrocorticographic activity. *Journal of Neuroscience*, 29(33):10203–10214, 2009.

[MLS+07] 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.

[MM12] Antonello Maruotti and Francesca Martella. Clustering multivariate longitudinal data: hidden markov of factor analyzers. In *46th Scientific Meeting of the Italian Statistical Society, Rome, Italy*, 2012.

[MVD97] Xiao-Li Meng and David Van Dyk. The em algorithm—an old folk-song sung to a fast new tune. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(3):511–567, 1997.

[Rab89] Lawrence R Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[Smy97] Padhraic Smyth. Clustering sequences with hidden markov models. *Advances in neural information processing systems*, pages 648–654, 1997.

[SR98] Lawrence K Saul and Mazin G Rahim. Modeling acoustic correlations by factor analysis. In *Advances in neural information processing systems*, pages 749–755, 1998.

[WCD+13] Wei Wang, Jennifer L Collinger, Alan D Degenhart, Elizabeth C Tyler-Kabara, Andrew B Schwartz, Daniel W Moran, Douglas J Weber, Brian Wodlinger, Ramana K Vinjamuri, Robin C Ashmore, John W Kelly, and Michael L Boninger. An electrocorticographic brain interface in an individual with tetraplegia. *PloS one*, 8(2):e55344, 2013.

[WOO+16] 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.

[YCS+09] Byron M Yu, John P Cunningham, Gopal Santhanam, Stephen I Ryu, Krishna V Shenoy, and Maneesh Sahani. Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. In *Advances in neural information processing systems*, pages 1881–1888, 2009.