ESTIMATING DIRECTED CONNECTIVITY VIA AUTOREGRESSIVE HIDDEN MARKOV MODELS

AMIRHOSSEIN KHALILIAN*

1. Introduction. Brain connectivity refers to several aspects of organization between different brain regions and can be studied using a broad range of analysis approaches. Three fundamental approaches to brain connectivity analysis have been studied in the literature: anatomical, functional, and directed connectivity.

Anatomical connectivity studies the physical or structural connections linking distinct units within a nervous system. Only invasive tracing studies are able to unanimously demonstrate these type of connections. *Functional connectivity*, on the other hand, is a fundamentally statistical concept. This type of connectivity studies the statistical dependencies between distributed and often spatially remote neuronal units. Measuring techniques such as correlation or covariance, spectral coherence or phase-locking can be utilized in this type of study. *Directed connectivity* describes the directional effect of one neural unit over another and in a sense is the union of structural and functional connectivity. This can also be referred to as discovering causal relations between distinct neural systems.

Here we focus on study of directed connectivity. In general the study of directed connectivity (or in fact any parametric causal inference) consists of application of three consecutive steps:

- model specification,
- model identification, and
- model inference.

Therefor in order to study the directed connectivity, we first need to specify a class of models represented with a set of parameters. We then identify the best model describing our data by fitting the model to the data and optimizing for the model parameters. Finally, we will leverage the characteristics of the identified model to infer the connectivity between different regions. Note that these steps are the major paradigm of any data-based causal inference such as Dynamic Causal Modeling (DCM) and Granger Causal Modeling (GCM).

In this study we investigate a class of models know as Autoregressive Hidden Markov Models (ARHMMs). Note that we are interested in analysing brain signals recorded while the subject is performing a task. It is safe to assume that the state of the brain changes at different parts of the task as time progresses. For instance, take *auditory repetition* as a task where the subject listens to a word and is asked to repeat it. One can assume four major states while performing the task: resting, perception, pre-articulation, and articulation. It is safe to assume that the switching between these states depends on the specific word used in each trial. Consequently, it is beneficial to leverage a class of models that can represent such behavior. By using ARHMM as our model class, we can learn the progression of the states in each trail as a Markov model in an unsupervised fashion.

When choosing a class of models for analysis it is crucial to select a class of models that are appropriate for future analysis and inference. Our ultimate goal is to infer the directed connectivity from the fitted model. ARHMMs have the added benefit that they present each brain-state by an AR model which makes the connectivity

*ECE Department, Tandon School of Engineering, New York University, (akg404@nyu.edu).

analysis possible via the relation of Granger causality and AR models.

In the following sections we will first briefly discuss the different notions of causality. Then, we introduce the AR and ARHMMs models. Additionally, we discuss the partial directed coherence which is a measure of Granger causality based on the AR coefficients. Finally, we provide experimental results on ECoG recorded signals.

2. Different notions of causality. The notion of causality is an epistemological concept and can be interpreted in different ways. It is beneficial to briefly review the common definitions of causality in the literature. Note that, the problem of tracking down the cause for a phenomenon, is a common yet complex question. A naive interpretation may suggest equating the causality with high values of correlation (i.e. degree of correlation between two variables or two time series). Although simple, this approach does not address the cause-and-effect relation and complexities associated with it. In order to address the complexities associated with understanding of causality, rigorous ways to approach this problem have been developed in different fields of science. The major fields dealing with such questions include econometrics, dynamic systems, and information theory to name a few.

Here, we are interested in causal inference over time series data, and thus we will be dealing with stochastic processes and causal relations between them. Our main objective is to review the major definitions and tools developed for causal inference especially in the field of neuroscience. More details can be found in review texts in the literature [5, 11, 13]. There are two distinct properties with practical relevance that causality can generally be defined by:

- *temporal precedence: cause precedes the effect;* and
- *physical influence: perturbation of the cause changes the effect* [13].

The temporal precedence is mainly considered in information theoretic definitions of causality such as Granger causal modeling (GCM). Whereas, physical influence is the back-bone of the methods from control theory such as dynamic causal modeling (DCM).

2.1. Granger causality. Granger causality is a popular method for defining and inferring causal relations in time-series data and was operationalized by Granger for autoregressive models [7]. In simple terms, Granger causality (G-causality) is based on temporal precedence and predictability. In formal terms, let X and Y be stationary random processes. Additionally, denote:

- all the information in the universe till time i with $\mathcal{U}_i = \{U_{i-1}, \cdots, U_{-\infty}\}$,
- all the information in X till time i with $\mathcal{X}_i = \{X_{i-1}, \cdots, X_{-\infty}\}$,
- the variance of the residual of predicting Y_i using $\mathcal{U}_i$ with $\sigma^2\left(Y_i|\mathcal{U}_i\right)$.

Then we can define X G-causes Y iff $\sigma^2\left(Y_i|\mathcal{U}_i\right) < \sigma^2\left(Y_i|\mathcal{U}_i/\mathcal{X}_i\right)$. Note that the required access to all the information of the universe is unrealistic. In practice, we can replace U with a limited set of observed time series.

2.2. Dynamic causal modeling. Physical influence speaks to the notion of intervention and control, and is the basis for the DCM type of causality [13]. In DCM we physically act upon (e.g., fix) the activity at one node and effectively remove any other physical influence this node receives. This means that inferences based on the effects of an intervention are somewhat different in nature from those based on purely observational effects and require proper experimental setup design and probabilistic calculus. On the other hand, in GCM the observations are the center of study. The focus of this article is on Granger type of causality.

2.3. A Tangential Discussion about a Causal Hierarchy. In this section, we briefly discus a topic slightly tangential but related to our main discussion and introduce a three-level hierarchy that arises from the theory of causal models. A more extensive discussion is presented in the technical report by Pearl [10]. This hierarchy classifies the causal information depending on the type of questions that each class is capable of answering. Table 1 gives a brief summary of each class in this hierarchy [10].

TABLE 1
A three-level hierarchy of causality from causal model theory [10]

Level	Mathematical Symbol	Typical Questions
1- Association	$P(y\|x)$	How would seeing X change my belief in Y?
2- Intervention	$P(y\|\text{do}(x), z)$	What if I do X?
3- Counterfactuals	$P(y_x\|x', y')$	What if I acted differently?

The first level is called "Association" since it purely involves statistical relations observed by the data and the questions at this level require no causal information. This level is characterized by conditional probability, i.e. the probability of event $Y = y$ given that we observed $X = x$. Bayesian networks or other machine learning tools are effective in performing such evidential computations from the data.

The second level, called "Intervention", entails not only observing what is, but also changing what we see. The questions in this level have the form "What will happen if we do X?" and because such questions cannot be answered purely by the data, interventions are placed higher than associations in the hierarchy. In terms of probability, this level is characterized as the probability of event $Y = y$ given that we intervene and set the value of $X = x$ and observe the event $Z = z$. Causal Bayesian networks and randomized trials are tools that can be utilized for answering such questions.

The last level of hierarchy, named "Counterfactuals", deals with models that answer the queries of the form "What would happen had we done X?". Note that such questions encompass the interventional and associational queries but not the other way around. Expressions of the type "the probability of that the event $Y = y$ would be observed had X been x, given that we actually observed $X = x'$ and $Y = y'$." characterize the Counterfactuals. These queries can only be answered when we possess functional or structural equation models.

This hierarchy, although tangential to our main discussion, illustrates a formal restriction of causal modeling. When dealing with cause-and-effect relationships, one needs to be mindful of the properties of the question being answered and choose the proper tools to define and solve the problem. Otherwise, however novel and sophisticated, the tools from the lower levels of the hierarchy cannot answer the queries related to higher levels.

3. Model specification and identification. In this section, we first introduce the autoregressive (AR) model and then extend it to the autoregressive hidden Markov model (ARHMM).

3.1. Autoregressive Model. Autoregressive processes are random processes with a specific temporal structure. In these models, the signal at each time point, $x(t) \in \mathbb{R}^n$ is a linear combination of the signal at previous time-points and a random innovation $\epsilon(t) \sim \mathcal{N}(0, I)$. The dynamics of the system is described by a tensor of AR coefficients $A = [A_p]_{p=1}^{P} \in \mathbb{R}^{P \times n \times n}$ (an $n \times n$ matrix for each time lag p) and a

covariance matrix Q. The dynamics of the system for an AR model of order P can be described as

$$x(t) = \sum_{p=1}^{p} A_p x(t-p) + Q^{\frac{1}{2}}\epsilon(t), \quad 1 \le t \le T. \tag{3.1}$$

Consequently, the parameters of the AR models are $\Theta = \{A, Q\}$.

In order to estimate the parameters of the AR model, equation (3.1) can be written as

$$x(t) + A_1 x(t-1) + \cdots + A_P x(t-P) = e(t)$$

where $e(t)$ is a zero-mean uncorrelated noise vector with covariance matrix Q. By multiplying with $x^T(t-k)$ for $k = 1, \cdots, P$ and taking expectation of both sides, we get to the Yule-Walker equations

$$R_{(-k)} + A_1 R_{(-k+1)} + \cdots + A_P R_{(-k+P)} = 0 \text{ where } R_k = x(t)x^T(t+k). \tag{3.2}$$

We estimate the covariance matrix R_k by

$$\widehat{R}_k = \frac{1}{T-k}\sum_{t=1}^{T} x(t)x^T(t+k),$$

and we average over trials if multiple trials are available. Note that (3.2) contains Pn^2 equations and the same number of unknown model parameters. Although one can simply solve these equations to obtain the model coefficients, the Levinson, Wiggins, Robinson (LWR) algorithm is a more robust solution procedure [9]. Note that the covariance matrix Q is estimated as a byproduct of the LWR algorithm [3]. For the estimated AR model to be stable, the roots of the characteristic polynomial

$$\det(\alpha^P I + \alpha^{P-1} A_1 + \cdots + \alpha A_{(P-1)} + A_P) = 0$$

must satisfy $|\alpha| < 1$ or equivalently, the largest eigenvalue of the companion matrix, F, must be smaller than one.

$$F = \begin{bmatrix} A_1 & A_2 & \cdots & A_P \\ I_n & 0 & \cdots & 0 \\ \vdots & \ddots & \cdots & \vdots \\ 0 & \cdots & I_n & 0 \end{bmatrix}$$

Note that the AR process is a linear model and has the inherent assumption that the signal x can be modeled with an stationary process. Ding et al. propose to model short windows of signal with separate AR models to overcome this challenge [3].

3.2. Autoregressive Hidden Markov Models. The AR model is linear and is not well suited for describing brain activity over a long period of time. Note that we are interested in analysing the brain activity while performing a task, and it is safe to assume that the state of the brain changes as time progresses. We ideally want to utilize a model that is rich enough to capture this state behavior.

Hidden Markov Model (HMM) is a latent state representation that describes the observed signals as a consequence of an unobserved latent state. The probability of occurrence of a state is modeled via a Morkov process. Let s_t denote the state of the system at each time point t, and $s_t \in \{1, \cdots, S\}$ for a total of S states. As a result,

the transition probability from state i at time t to state j at time $t+1$ can be given as $\Phi_{i,j} = \Pr(s_{t+1} = j | s_t = i)$.

Autoregressive Hidden Markov Model combines AR stochastic dynamics with HMMs such that each latent state indicates a different AR process [6]. As a result, for each state a different AR process with state-specific dynamics and noise covariance are estimated. Note that the switching between states is controlled by a Markov process and is estimated in an unsupervised manner. Additionally, the switching of states makes the ARHMM effectively nonlinear. A pictorial representation of the ARHMM model is shown in figure 1. Application of ARHMM for different types of brain activity has been studied in the literature [2, 12].

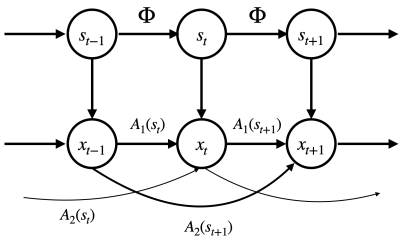


FIG. 1. *A pictorial representation of the ARHMM.*

To formally introduce the ARHMM, let s_t denote the state of the system at each time point t, and $s_t \in \{1, \cdots, S\}$ for a total of S states. The transition matrix of the states is $\Phi = [\Phi_{i,j}] = [\Pr(s_{t+1} = j | s_t = i)] \in \mathbb{R}^{S \times S}$ with initial state probability $\phi_0 \in \mathbb{R}^S$. For each state s we have the AR coefficients $A_p(s)$ and the noise covariance matrix $Q(s)$ where $p = 1, \cdots, P$ for AR models of order P. Then, the ARHMM model can be written as

$$x(t) = \sum_{p=1}^{p} A_p(s_t) x(t-p) + Q^{\frac{1}{2}}(s_t)\epsilon(t), \quad 1 \le t \le T. \tag{3.3}$$

Consequently, the parameters that need to be estimated in the ARHMM can be described as $\Theta = \{\Phi, \phi_0, A(s), Q(s) : \forall s = 1, \cdots, S\}$.

The parameters of ARHMM can be estimated via the expectation maximization (EM) algorithm [4]. The key steps of the EM algorithm are Expectation step (E-step) which computes the expectation of the likelihood function by including unobserved data as if they were observed, and Maximization step (M-step) which updates the estimate of model parameters by maximizing the expected likelihood function computed in E-step. The Likelihood function after each E-step can be written as

$$L(\Theta) = \mathbb{E}\left\{\log \Pr(x_{1:T}, s_{1:t} | \Theta) | x_{1:T}, \Theta^{\text{Old}}\right\}.$$

The M-step estimates the model parameters Θ by computing $\arg\max_\Theta L(\Theta)$.

4. Model inference: Estimating Granger Causality from AR model. In this section we describe the method to estimate the directed connectivity from the fitted AR coefficients. Directed coherence (DC), discrete transfer function (DTF), and Granger causality test (GCT) are examples of the possible avenues for estimating the Granger causality from the AR model. In this study, we use the notion of the partial directed coherence which is a frequency-domain approach to describing the relationships (direction of information flow) between multivariate time series [1].

4.1. Partial Directed Coherence. n order to analyze the Granger causality in terms of the AR model and also provide a frequency domain picture for Granger causality, Baccala et al. defined the partial directed coherence (PDC) [1]. Although one may statistically test for the hypothesis $A_{ij} = 0$, a frequency domain picture is missing for most of such tests. The PDC factor is defined as

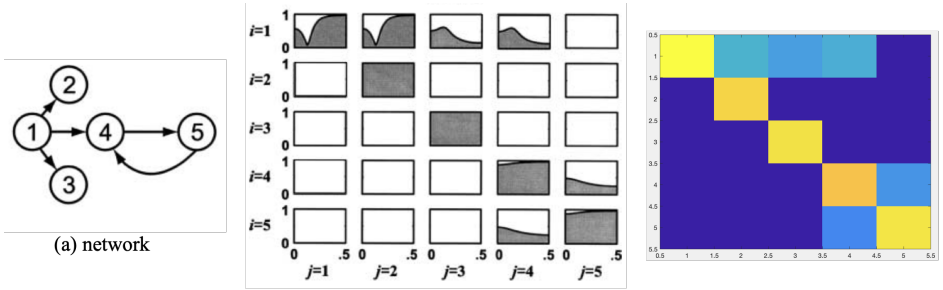
$$\pi_{ij}(f) = \frac{\bar{A}_{ij}(f)}{\sqrt{\bar{A}^H_{\cdot j}(f)\bar{A}_{\cdot j}(f)}} \tag{4.1}$$

where

$$\bar{A}_{ij}(f) = \begin{cases} 1 - \sum_{\tau=1}^{p} A_{ij}(\tau)\exp\left(-2\pi i f \tau\right), & \text{if} \quad i = j \\ -\sum_{\tau=1}^{p} A_{ij}(\tau)\exp\left(-2\pi i f \tau\right) & \text{otherwise} \end{cases}$$

Some properties of PDC are $0 \leq |\pi_{ij}(f)|^2 \leq 1$ and $\sum_{i=1}^{n} |\pi_{ij}(f)|^2 = 1 \quad \forall 1 \leq j \leq n$. As a result, PDC ranks the relative interaction strengths with respect to a given signal source. Note that the PDC matrix measures the Granger causality from AR coefficients. As a result, PDC can be calculated for each state of the ARHMM separately based on the estimated AR coefficients. An example from [1] showing the PDC for a network described by following equations is presented in figure 2.

$$\begin{cases} x_1(t) = 0.95\sqrt{2}\,x_1(t-1) - 0.9025\,x_1(t-2) + \epsilon_1(t) \\ x_2(t) = 0.5\,x_1(t-1) + \epsilon_2(t) \\ x_3(t) = -0.4\,x_1(t-3) + \epsilon_3(t) \\ x_4(t) = -0.5\,x_1(t-2) + 0.25\sqrt{2}\,x_4(t-1) + 0.25\sqrt{2}\,x_5(t-1) + \epsilon_4(t) \\ x_4(t) = -0.25\sqrt{2}\,x_4(t-1) + 0.25\sqrt{2}\,x_5(t-1) + \epsilon_5(t) \end{cases}$$



(a) network (b) PDC in frequency (c) PDC coefficient

FIG. 2. *An example of the PDC for a network [1]*

5. Experimental results. In this section, we provide a few experimental results showing the estimated states and connectivity matrices for a subject performing the auditory repetition task. We will show results for the same trial data when locked to perception and locked to production.

5.1. Experimental setup. the auditory repetition task involves a subject hearing a word and then repeating it. The ECoG signal is recorded during the task via a grid of electrodes covering the cortical regions related to language processing. ECoG

signals were preprocessed with high-gamma band-pass filter (70-150 Hz). The envelope of the filtered signal was then extracted by a Hilbert Huang transform. We normalize the signal from each electrode by its mean and standard deviation. Finally, the signal was downsampled to 200 Hz. We identified twenty electrodes that are active during the task and are in the regions related to the auditory task. The selected electrodes are shown in figure 3.

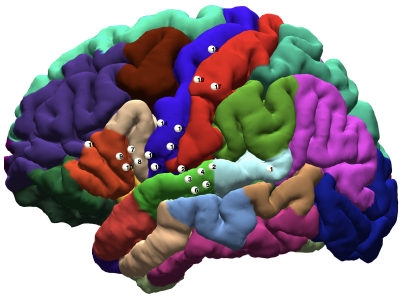


FIG. 3. *Selected electrodes depicted on the annotated brain.*

For each experiment, as an initialization, we fit an AR model to windows of 100ms with 50ms overlap. Then we calculate the PDC for each window and cluster the windows based on the PDCs using k-means clustering algorithm. The AR parameters for each state in ARHMM are initialized by the average of the AR parameters in the corresponding cluster.

5.2. ARHMM results when locked to *perception* period. In this section we present the results when we select windows of the recorded signal such that each window starts 60ms before the stimulus is presented at each trail and the window length is 2 second. The signal for the selected electrodes is shown in figure 4.

The resulting states and PDC matrices from the ARHMM model with S=5, P=3 are shown in figure 5. Figure 5(a) shows the estimated states per trial per time point on the left and the probability of the estimated state on the right. The red lines show the start and end of stimulus and production part as depicted on the time-line. As we can observe, s=1 corresponds to a resting state and the PDC for this state does not show any activity as expected. The next state is s=3 which starts shortly after the stimulus starts. The PDC for this state shows connectivity mostly from the STG region to IFG and precentral and postcentral (motor and sensory) regions. The next, s=2, is the pre-articulation state. The PDC for this states shows activity within IFG and toward sensory and motor regions. Finally, states s=4 and s=5 correspond to articulation. The PDC for these states show activity from motor and sensory regions back to STG. The recovered states and observed relations follow the expected behavior from the literature.

5.3. ARHMM results when locked to *production* period. In this section we present the results when we select windows of the recorded signal such that each window is centered at when the speech production starts at each trail and the window length is 2 second. The signal for the selected electrodes is shown in figure 6.

The resulting states and PDC matrices from the ARHMM model with S=5, P=3 are shown in figure 5. Figure 5(a) shows the estimated states per trial per time point on the left and the probability of the estimated state on the right. The red lines show the start and end of stimulus and production part as depicted on the time-line.

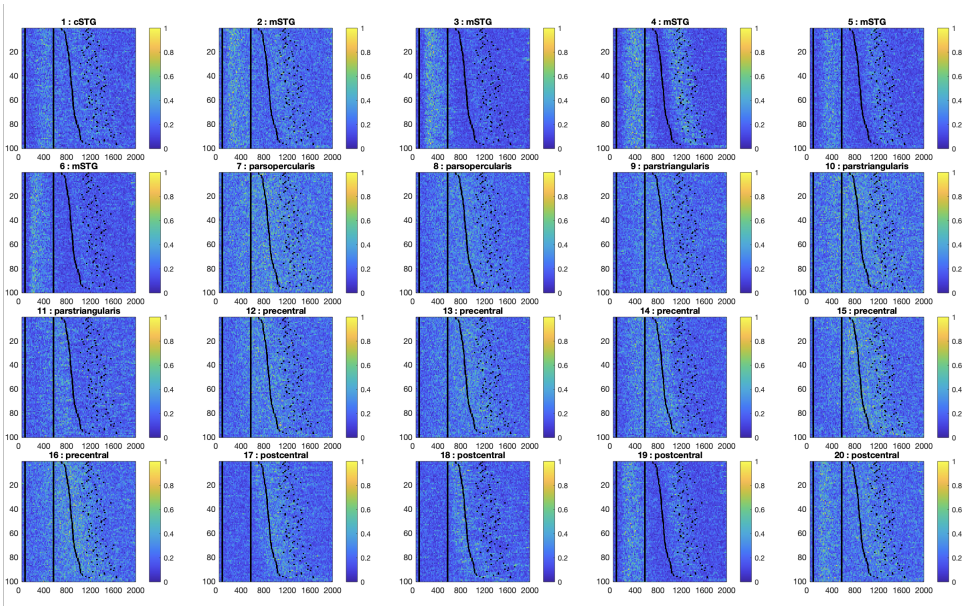


FIG. 4. *Signal in each electrode when locked to perception period. The x-axis on each plot shows the time and the y-axis shows the trial number. The black lines show start and end of stimulus and start and end of production, respectively.*

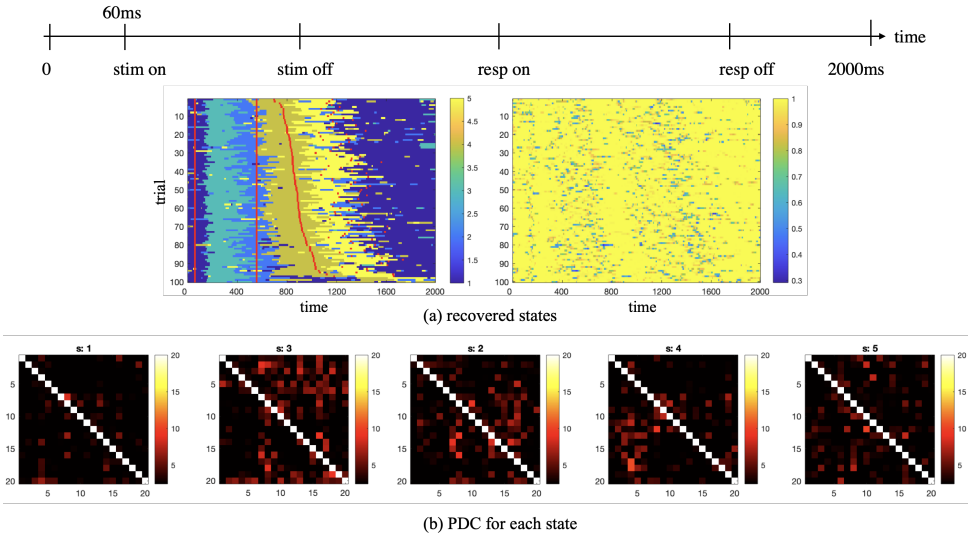


FIG. 5. *Result of the ARHMM for signal when locked to perception.*

The recovered states in this case and their corresponding PDC matrices show similar behavior as the case when locked to perception. The flow of information follows similar pattern that further enforces the results of the ARHMM.

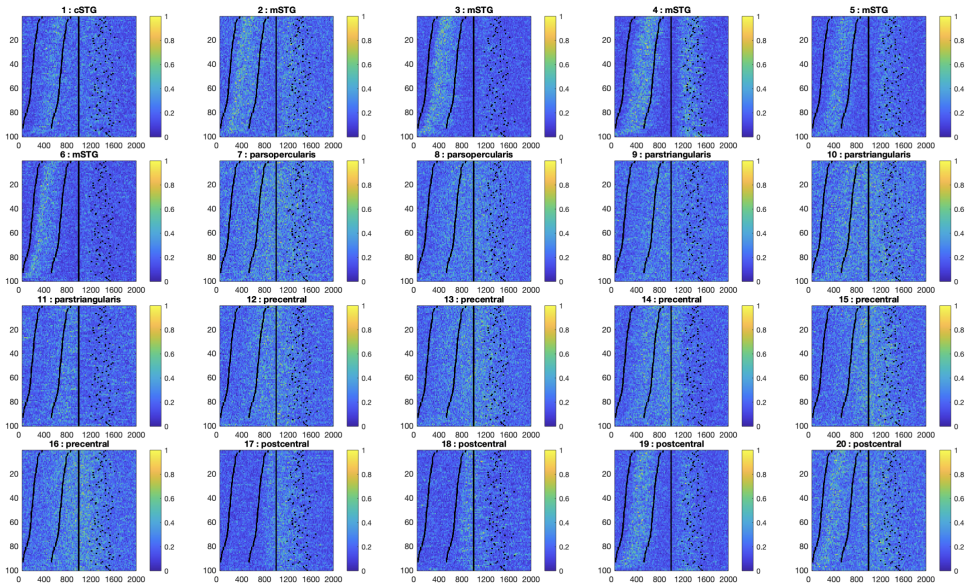


FIG. 6. *Signal in each electrode when locked to production period. The x-axis on each plot shows the time and the y-axis shows the trial number. The black lines show start and end of stimulus and start and end of production, respectively.*

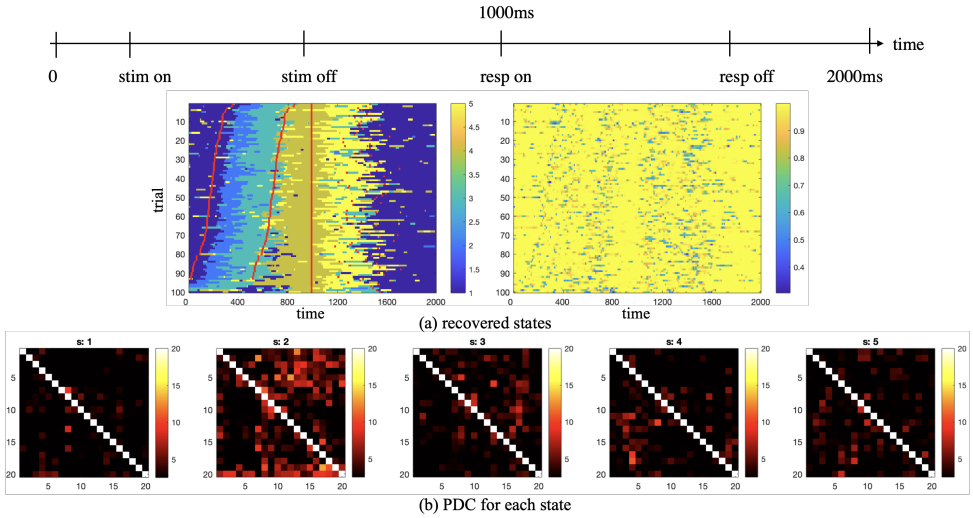


FIG. 7. *Result of the ARHMM for signal when locked to production period.*

6. Discussion and Conclusion. In this article we reviewed some of the techniques for estimating the directed connectivity between different cortical regions from ECoG recordings. We utilized the ARHMM method to model the signal in time as a set of AR-processes that are switched via a hidden state variable. This allows us to estimate the brain-state in an unsupervised fashion and reduce the assumptions on the model. Additionally, the switching between states makes the ARHMM effectively non-linear and more suitable for describing brain activity. Furthermore, the

connectivity defined by Granger causality can be estimated via the notion of partial directed coherence from the AR coefficients at each state. Additionally, partial directed coherence gives a frequency domain picture of the directed connectivity that can be beneficial for further analysis.

In future work we aim to further extend the developed framework and investigate the results of ARHMM for different tasks and between different subjects. Note that an inherent problem of the ARHMM model and more specifically solving for the model parameters via expectation maximization is the computational instability of this algorithm as the number of electrodes increases. Recently, Linderman et al. proposed a Bayesian framework for learning and inference of a class of models closely related to ARHMM [8]. They leverage the recent Poly-gamma auxiliary variable techniques and develop algorithmic solutions in the framework of Bayesian learning that are scalable, fast, and efficient. We aim to study such techniques to develop more stable algorithms that allow for solve for more cortical nodes and investigate the connectivity in finer details.

REFERENCES

[1] L. A. Baccalá and K. Sameshima, *Partial directed coherence: a new concept in neural structure determination*, Biological cybernetics, 84 (2001), pp. 463–474.

[2] J. Chiang, Z. J. Wang, and M. J. McKeown, *A hidden Markov, multivariate autoregressive (HMM-mAR) network framework for analysis of surface EMG (sEMG) data*, IEEE Transactions on Signal Processing, 56 (2008), pp. 4069–4081.

[3] M. Ding, S. L. Bressler, W. Yang, and H. Liang, *Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: data preprocessing, model validation, and variability assessment*, Biological cybernetics, 83 (2000), pp. 35–45.

[4] C. B. Do and S. Batzoglou, *What is the expectation maximization algorithm?*, Nature biotechnology, 26 (2008), pp. 897–899.

[5] M. Eichler, *Causal inference in time series analysis*, Wiley Online Library, 2012.

[6] E. Fox, E. B. Sudderth, M. I. Jordan, and A. S. Willsky, *Nonparametric bayesian learning of switching linear dynamical systems*, in Advances in neural information processing systems, 2009, pp. 457–464.

[7] C. W. Granger, *Investigating causal relations by econometric models and cross-spectral methods*, Econometrica: journal of the Econometric Society, (1969), pp. 424–438.

[8] S. Linderman, M. Johnson, A. Miller, R. Adams, D. Blei, and L. Paninski, *Bayesian learning and inference in recurrent switching linear dynamical systems*, in Artificial Intelligence and Statistics, 2017, pp. 914–922.

[9] M. Morf, A. Vieira, D. T. Lee, and T. Kailath, *Recursive multichannel maximum entropy spectral estimation*, IEEE Transactions on Geoscience Electronics, 16 (1978), pp. 85–94.

[10] J. Pearl, *The seven tools of causal inference, with reflections on machine learning*, Communications of the ACM, 62 (2019), pp. 54–60.

[11] J. Runge, *Detecting and quantifying causality from time series of complex systems*, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät, 2014.

[12] A. G. Saravani, K. J. Forseth, N. Tandon, and X. Pitkow, *Dynamic brain interactions during picture naming*, eNeuro, 6 (2019).

[13] P. A. Valdes-Sosa, A. Roebroeck, J. Daunizeau, and K. Friston, *Effective connectivity: influence, causality and biophysical modeling*, Neuroimage, 58 (2011), pp. 339–361.