Neuroscience Research 81–82 (2014) 69–77

Contents lists available at ScienceDirect

Neuroscience Research

journal homepage: www.elsevier.com/locate/neures

# An artificial network model for estimating the network structure underlying partially observed neuronal signals

Misako Komatsu[a,b], Jun Namikawa[a], Zenas C. Chao[a], Yasuo Nagasaka[a], Naotaka Fujii[a], Kiyohiko Nakamura[b], Jun Tani[c,*]

[a] RIKEN Brain Science Institute, 2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan
[b] Graduate School of Science and Engineering, Tokyo Institute of Technology, 4259 Nagatsuta, Midori-ku, Yokohama 226-8502, Japan
[c] Department of Electrical Engineering, Korea Advanced Institute of Science and Technology, 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Republic of Korea

ARTICLE INFO

*Article history:*
Received 10 October 2013
Received in revised form 5 February 2014
Accepted 5 February 2014
Available online 14 February 2014

*Keywords:*
Functional connectivity
Directed interaction
ECoG
Causality
Intracranial EEG

ABSTRACT

Many previous studies have proposed methods for quantifying neuronal interactions. However, these methods evaluated the interactions between recorded signals in an isolated network. In this study, we present a novel approach for estimating interactions between observed neuronal signals by theorizing that those signals are observed from only a part of the network that also includes unobserved structures. We propose a variant of the recurrent network model that consists of both observable and unobservable units. The observable units represent recorded neuronal activity, and the unobservable units are introduced to represent activity from unobserved structures in the network. The network structures are characterized by connective weights, i.e., the interaction intensities between individual units, which are estimated from recorded signals. We applied this model to multi-channel brain signals recorded from monkeys, and obtained robust network structures with physiological relevance. Furthermore, the network exhibited common features that portrayed cortical dynamics as inversely correlated interactions between excitatory and inhibitory populations of neurons, which are consistent with the previous view of cortical local circuits. Our results suggest that the novel concept of incorporating an unobserved structure into network estimations has theoretical advantages and could provide insights into brain dynamics beyond what can be directly observed.


## 1. Introduction

Recent methodological developments in the simultaneous recording of multi-dimensional brain signals have increased our understanding of how these signals interact with each other, and how information is processed in the brain. To analyze the relationships between simultaneously recorded neurons, bivariate statistical dependencies can be computed using measures such as cross-correlations, coherence, and phase synchronization (e.g., Griffith and Horn, 1963; Bressler et al., 1993; and Cobb et al., 1995, respectively). While these methods can be effective in identifying potential functional connections between pairs of neurons, or between brain regions, they are unable to capture any directional (or causal) interactions. To address this issue, researchers have recently begun to develop alternative causal measures. One of the most widely used methods is the Granger causality, which was originally developed for use with social and economic systems (Granger, 1969). On the basis of a time-lagged linear regression analysis, Granger causality captures information about the future state of 1 variable by taking into account the past states of another variable. Transfer entropy is a related measure based on an information theory that is also based on linear interactions (Schreiber, 2000). As another causal measure, transfer entropy detects directional interactions between 2 variables by considering the effects of the state of 1 variable on the state transition probabilities of another. However, these causal measures cannot detect whether the causality is due to direct or indirect interactions, or to a common external influence, such as a shared input (Kaminski et al., 2001).

More recently, these measures have been extended in a multivariate manner, and applied to the study of causal relationships in neuronal networks (e.g., Hesse et al., 2003; Smith et al., 2006; Quinn et al., 2011). These studies have attempted to determine the directional influences between any given pair of signals from

* Corresponding author. Tel.: +82 42 350 7428.
*E-mail addresses:* mkomatsu@brain.riken.jp (M. Komatsu), jnamika@gmail.com (J. Namikawa), zenas.c.chao@gmail.com (Z.C. Chao), nyasuo@brain.riken.jp (Y. Nagasaka), na@brain.riken.jp (N. Fujii), nakamura@dis.titech.ac.jp (K. Nakamura), tani1216jp@gmail.com (J. Tani).

http://dx.doi.org/10.1016/j.neures.2014.02.005
0168-0102/

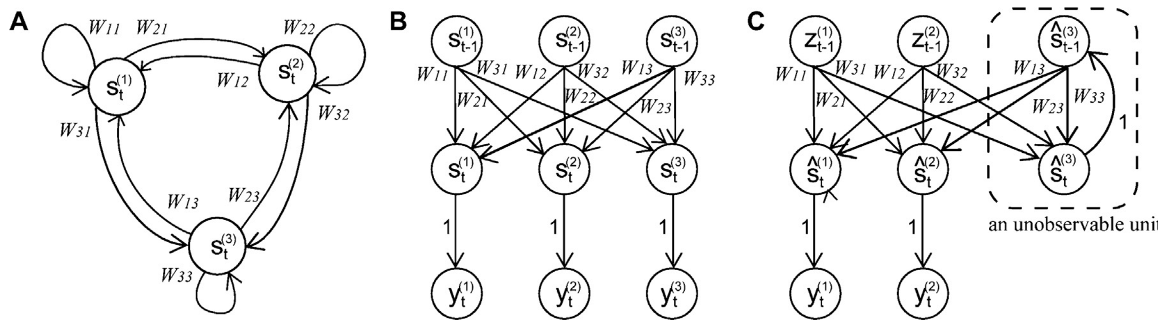


**Fig. 1.** Introduction of the proposed network. (A) An example of a recurrent network consisting of 3 units. These units interact with each other through their connective weights $W_{ij}$, and update their outputs $\hat{s}_t^{(i)}$. (B) Focus on the time domain of the recurrent network. The network contains the input layer, consisting of the previous outputs of each unit at time $t-1$, and the output layer, consisting of the current outputs at time $t$. (C) The partially observable network. The inputs and outputs of the network are provided only from observed signals. The middle layer consists of both observable and unobservable units.

multi-channel recordings, and have provided insights into the causal relationships between those signals. It is important to note that to date, such studies have not inferred hidden neural structures that underlie those causal interactions (Friston, 2009), although signals have only been observed from a part of the system. In order to obtain a more complete picture of information processing in the brain from only partially observed activity, we need to account for the influence of neural structures that are not directly observed. To this end, we aimed to reveal the influence of unobserved neural structures, as well as the multivariate causal interactions between recorded neuronal signals.

Hidden Markov models (HMMs) constitute one of the more widely used models, and explicitly include such unobserved structures of data. Some studies have attempted to infer the hidden structure of HMM (Viterbi, 1967; Baum et al., 1970). However, because conventional HMMs use symbol-level abstraction, they are not ideal for attempting to describe the neural basis of the observed signals. In the present study, we use a rate-coding recurrent network. In the network, the units generate real-valued time series of their outputs through the interactions. Those outputs represent the average firings of neuronal ensembles. Furthermore, the network is capable of a variety of dynamic behaviors, including chaotic dynamics (Funahashi and Nakamura, 1993; Beer, 1995). Recently, some studies have successfully applied the neural network to macroscopic behavior of a neuronal population (Yamashita et al., 2008, 2011). The recurrent network of rate-coding units is thus considered able to emulate the characteristic features of macroscopic mechanisms of biological neural systems.

For estimating network structures underlying partially observed signals, we propose a variant of the recurrent network model. The model includes a structure that preserves the influence from unobserved structures underlying observed signals. The model also acquires network parameters through iterative training methods, using only the observed signals. Then, we applied this model to multi-dimensional brain signals recorded from monkeys, and demonstrated that the model could reconstruct a robust network structure with physiological plausibility. Further, we discuss functional interpretations of the network. This study extends the work presented in Proceedings of the 2012 International Joint Conference on Neural Networks (Komatsu et al., 2012).

## 2. Methods

### 2.1. Network model

In this section, we first propose a variant of a three-layer recurrent network. The network explicitly implements the hidden structures of observed signals. Then, we explain how the structure of the network was inferred from the observed signals.

#### 2.1.1. Partially observable network

First of all, we consider a recurrent network as a biological model of macroscopic neural networks (Fig. 1A). The units update their states through their interactions. Let $c_t^{(i)}$, $s_t^{(i)}$, and $W_{ij}$ denote the state and output of unit $i$ at time $t$, and the connective weight indicating the strength of interaction from unit $j$ to unit $i$, respectively. Then, the state and output are updated according to the following equations:

$$c_t^{(i)} = \left(1 - \frac{1}{\tau^{(i)}}\right) c_{t-1}^{(i)} + \frac{1}{\tau^{(i)}} \sum_{j=1}^{N} W_{ij} s_{t-1}^{(j)}, \tag{1}$$

$$s_t^{(i)} = \tanh(c_t^{(i)} + h^{(i)}), \quad i = 1, 2, \ldots, N, \tag{2}$$

where $N$, $\tau^{(i)}$, and $h^{(i)}$ are the number of units, the time constant of unit $i$, and the bias of unit $i$, respectively. Eq. (1) means that the internal state of unit $i$ at time $t$ is given as an interpolation between the previous state at time $t-1$ and the linear summation of the outputs of all units at time $t-1$. Eq. (2) implies an assumption that each output has some limited values.

Those updates of the states of the units can be described as a multi-input-multi-output network (Fig. 1B). The network has the input layer as the previous outputs of each unit at time $t-1$, and the output layer as the current outputs at time $t$: that is

$$y_t^{(i)} = s_t^{(i)}, \quad i = 1, 2, \ldots, N, \tag{3}$$

where $y_t^{(i)}$ are the outputs of units of the output layer.

Here, we suppose that not all such units within the network are observable. Fig. 1A provides an illustrative example; suppose that the outputs of units 1 and 2 are observable, and that the output of unit 3 is not. In this case, the multi-input–multi-output network shown in Fig. 1B is transformed to the network shown in Fig. 1C. Fig. 1B and C represents the same network. The difference between them is that the inputs and outputs of the network in Fig. 1C are provided only by observed signals, while the middle layer consists of both observable and unobservable units. The observed signals at time $t$ are denoted by $z_t^{(k)}$, $k = 1, 2, \ldots$, CH, where CH is the total number of recorded units. The outputs of observable and unobservable units at time $t$ are also denoted by $\hat{s}_t^{(i)}$, $i = 1, 2, \ldots, N$, where $N$ is the total number of units. In the network, the state of the unobservable unit $\hat{s}_{t-1}^{(3)}$ is provided through the connective weights $W_{13}$, $W_{23}$, and $W_{33}$ in the middle layer. In this study, we refer to this variant of recurrent networks as partially observable networks (PONs). In PONs, there are connections from the input layer to the middle layer, from the unobservable units to both the observable and unobservable units, and the observable units have connections only to the output layer, fixed with a connective weight of 1 (Fig. 1C). A similar architecture of three-layer recurrent networks was originally employed by Elman (1990) to represent time in connectionist