Neural Networks 28 (2012) 1–14

Contents lists available at SciVerse ScienceDirect

Neural Networks

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

A dynamical pattern recognition model of gamma activity in auditory cortex

M. Zavaglia[b], R.T. Canolty[c], T.M. Schofield[a], A.P. Leff[d], M. Ursino[b], R.T. Knight[c], W.D. Penny[a],*

[a] Wellcome Trust Centre for Neuroimaging, University College, London WC1N 3BG, UK
[b] Department of Electronics, Computer Science and Systems (DEIS), Via Venezia 52, 47023 Cesena, Italy
[c] Helen Wills Neuroscience Institute, University of California, Berkeley, CA, USA
[d] Institute of Neurology and Institute of Cognitive Neuroscience, UCL, 17 Queen Square, London WC1N 3AR, UK

ARTICLE INFO

Article history:
Received 11 April 2010
Received in revised form 20 December 2011
Accepted 21 December 2011

Keywords:
Gamma activity
Speech recognition
Synchronization
Transients
Coupled-oscillator
Bayesian estimation

ABSTRACT

This paper describes a dynamical process which serves both as a model of temporal pattern recognition in the brain and as a forward model of neuroimaging data. This process is considered at two separate levels of analysis: the algorithmic and implementation levels. At an algorithmic level, recognition is based on the use of Occurrence Time features. Using a speech digit database we show that for noisy recognition environments, these features rival standard cepstral coefficient features. At an implementation level, the model is defined using a Weakly Coupled Oscillator (WCO) framework and uses a transient synchronization mechanism to signal a recognition event. In a second set of experiments, we use the strength of the synchronization event to predict the high gamma (75–150 Hz) activity produced by the brain in response to word versus non-word stimuli. Quantitative model fits allow us to make inferences about parameters governing pattern recognition dynamics in the brain.

1. Introduction

Hopfield and Brody (2000, 2001) (HB) have proposed a model for how the brain might recognize spatiotemporal patterns, and have applied it to the problem of auditory word recognition. Their model is particularly appealing at two different levels of analysis (Marr & Poggio, 1976).

First, at an 'algorithmic' level the HB model uses a preprocessing stage comprising a bank of filters and a set of feature detectors which signal onsets, offsets and peak activities in different frequency ranges. This is broadly consistent with the physiology of the mammalian auditory system (Casseday, Fremouw, & Covey, 2002; Ghitza, 1986). The key aspect of their algorithm, however, is that the subsequent pattern recognition is based on the Occurrence Times (OTs) of features which provides a natural invariance to the speed at which a word is spoken.

Second, at an 'implementation' level the recognition of OTs is achieved using a transient synchronization mechanism. This phenomenon relies on a combination of three physiological processes acting in concert (i) spike rate adaptation, (ii) synaptic plasticity and (iii) neuronal synchronization. In the HB model synchronization arises via balanced excitation and inhibition (Tsodyks, Mitkov, & Sompolinsky, 1993) in a network of Integrate and Fire (IF) cells. Together these mechanisms provide a burst of gamma activity that corresponds to a 'recognition event'. This is particularly interesting to imaging neuroscientists as bursts of gamma activity (which we define here to be higher than 30 Hz in frequency) have been observed to accompany auditory word recognition (Canolty et al., 2007; Lutzenberger, Pulvermuller, & Birbaumer, 1994; Pulvermuller et al., 1996).

This paper draws heavily on the HB model and makes three new contributions to the literature. First, we consider the algorithmic level and use a speech database to assess the usefulness of OT features as compared to standard features used in Automatic Speech Recognition (ASR) that are based on cepstral coefficients (Rabiner & Juang, 1993). Both types of features (OT or cepstral) are then used as input to an identical pattern recognition module. This allows us to assess the usefulness of the features themselves independently of the utility of the pattern recognition process or its putative neurobiological implementation.

Second, we propose a more generic model of transient synchronization based on a Weakly Coupled Oscillator (WCO) framework (Hoppensteadt & Izhikevich, 1997). WCOs are a standard approach for studying synchronization dynamics (Hoppensteadt & Izhikevich, 1997) and can be derived by applying a phase reduction approach to neurophysiologically realistic neural (Gutkin, Ermentrout, & Reyes, 2005) or neural network (Brown, Moehlis, & Holmes, 2004) models. The only requirement is that the underlying neurons operate around a limit cycle and interact weakly (Brown et al., 2004; Ermentrout & Kleinfeld, 2001; Hansel, Mato, & Meunier, 1995).

---

* Corresponding author.
E-mail address: w.penny@fil.ion.ucl.ac.uk (W.D. Penny).

0893-6080/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2011.12.007

This paper uses a WCO model of transient synchronization which we refer to as the WCO–TS model. As in the HB model, recognition is signalled by a transient synchronization event, and this synchronization is brought about by coupling feature detectors that have nonstationary, pattern-dependent frequency response profiles. However, the synchronization process itself is not implemented using balanced excitation and inhibition among IF cells as in Hopfield and Brody (2001), but is rather described at the level of phase dynamics. This allows us to be equivocal about the details of the neural circuits that generate the oscillations themselves. We see this as a benefit as there are currently a large number of possible candidates for the underlying processes (see next section).

Third, we show how the WCO–TS model can be directly fitted to neuroimaging data. This follows the example of 'Dynamic Causal Modelling' in which differential equation models of physiological processes are fitted to data and scored against each other using Bayesian inference (Friston, Harrison, & Penny, 2003; Girolami, 2008; Penny, Litvak, Fuentemilla, Duzel, & Friston, 2009; Penny, Stephan, Mechelli, & Friston, 2004). Specifically, we show how the WCO–TS model can be used as a forward model of gamma activity observed in Electrocorticographic (ECOG) data.

The paper is organized as follows. The following subsection briefly reviews the topics of gamma activity and network synchronization. Section 2.1 then describes the ECOG data and the spectral analysis methods used to find the underlying gamma burst associated with word recognition. This is based on previous work (Canolty et al., 2007). Section 2.2.6 then describes the WCO–TS model and how it is fitted to data. The results section reports on the efficacy of OT features as assessed using a spoken digit database, and on the use of WCO–TS as a forward model of ECOG data.

1.1. *Gamma activity and synchronization*

The phenomenon of gamma activity has received tremendous interest in imaging neuroscience. It initially rose to prominence with regard to the feature binding problem, whereby features of the same object that are represented in different brain regions must somehow be tied together to form a coherent whole. It was proposed that synchronization between the relevant regions at gamma frequency was just such a mechanism (Singer, 1999). There has since been a large amount of work in this area with reviews focusing on its role in large-scale integration (Varela, Lachaux, Rodriguez, & Martinerie, 2001), enhanced communication (Fries, 2005), attention and memory (Jensen, Kaiser, & Lachaux, 2007) and spike-timing dependent plasticity (Buzsaki, 2006). Gamma is also the single frequency band which most strongly predicts BOLD activity (Goense & Logothetis, 2008). We are therefore interested in gamma activity as it potentially provides a connection between computational and imaging neuroscience.

In the auditory domain several studies have found stronger (25–35 Hz) gamma responses to words as opposed to pseudo-words (Lutzenberger et al., 1994; Pulvermuller et al., 1996) and in the 60–70 Hz range to words as opposed to non-words (Eulitz et al., 1996). Additionally, Canolty et al. (2007) have found High Gamma (80–200 Hz) responses in ECOG recordings to words as opposed to non-words. Additionally, this High Gamma activity occurred sequentially over posterior Superior Temporal Gyrus (STG), mid STG, followed by Superior Temporal Sulcus (STS). This extends previous findings from fMRI (Binder et al., 2000) and provides evidence for a degree of seriality in word processing. It is this data set that we will analyse using the WCO–TS model.

The above neuroimaging results and related conceptual advances have motivated a number of theoretical models. For example, Shamir, Ghitza, Epstein, and Kopell (2009) have developed a neurophysiologically realistic model that shows how gamma oscillations can directly represent stimuli whose time scale is longer than a single gamma cycle, as is required for the representation of auditory words. Hopfield (2004) shows that subthreshold oscillations can be used to support a spike-time based code that leads to minimal interference with coexisting firing rate codes, and that subthreshold oscillations at gamma frequency may be important for encoding of speech. This principle has been developed by Ghitza (2007) who also propose that hierarchies of rhythms may be the mechanism by which the brain integrates information over multiple time scales during language processing.

We now turn to the issue of what is the physiological origin of gamma activity. As with most oscillatory phenomena in the brain, gamma is thought to arise from a combination of factors (i) a cell's intrinsic ability to oscillate, (ii) the presence of feedback connections among groups of excitatory and inhibitory neurons and (iii) the ability of networks of cells to either amplify or nullify certain oscillations. These factors are described in a recent comprehensive review (Wang, 2010). One mechanism for network amplification is the synchronization of cell activity.

The frequency of oscillations produced by single cells is determined primarily by the synaptic time constants and levels of driving input, with faster synapses and stronger inputs generally leading to higher frequency oscillations. These oscillations require that cells receive a tonic excitatory drive. When two cells are connected the resulting activity depends on whether the intervening interactions are fast or slow.

Mathematical studies of coupled oscillators show that for fast interactions, synchronization is most readily achieved using excitatory connections (Vreeswijk, Abbott, & Ermentrout, 1994). In the mammalian brain fast excitatory connections can be mediated by electrical synapses or gap junctions. These are found, for example, between pyramidal cells in hippocampus. In neural network models with tonic drive, gap junctions can lead to synchronized gamma activity (Pfeuty, Mato, Golomb, & Hansel, 2003). Traub, Schmitz, Jefferys, and Draguhn (1999), have shown using simulations that a network of pyramidal cells, electrically coupled through their axons, can generate High Gamma activity without chemical synapses.

If the interactions are slow then synchronization is most readily achieved using inhibitory connections. Chemical synapses with realistic rise times fall into this 'slow' category. For a pair of IF cells receiving tonic excitation, synchronization can be achieved using mutual inhibition (Vreeswijk et al., 1994). This result follows over to conductance-based models with large numbers of cells (Tiesinga & Jose, 2000; Wang & Buzsaki, 1996; White, Chow, Ritt, Soto-Trevino, & Kopell, 1998). These network models are referred to as Inhibitory Network Gamma (ING) oscillators (Bartos, Vida, & Jonas, 2007). ING oscillators have slow synapses and connections are weak. For these oscillations to impact on signals sent from a region they must recruit pyramidal cells which then in turn re-excite local interneurons. This results in so-called Pyramidal Inhibitory Network Gamma (PING) oscillators (Whittington, Traub, Kopell, Ermentrout, & Buhl, 2000).

A potential problem with ING/PING oscillators is that they are sensitive to parameter inhomogeneities between cells. If cells receive different input drives then synchronization can be destroyed (Wang & Buzsaki, 1996). Gamma oscillations that are resistant to such inhomogeneities, however, can be generated with ING oscillators having strong rather than weak synapses, fast rather than slow synapses, and with inhibition that is shunting (i.e., vetoing any excitatory input) rather than merely hyperpolarizing (Bartos et al., 2007). In mammalian neocortex the fastest synapses exist in the form of gap junctions between layer 4 inhibitory interneurons. These junctions promote synchronization without changing network frequency (Bartos et al., 2007).