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
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).

In the networks we have so far described both excitatory and inhibitory cells fire, approximately, on every gamma cycle so that the Local Field Potential (LFP) oscillation frequency is the same as the firing rate of the cells. Brunel and Hakim (1999) have investigated a different regime they call Weak Stochastic Synchronization (WSS) in which interneurons and pyramidal cells fire stochastically and during a small proportion of gamma cycles only. WSS can be brought about by combining strong synapses with noise. This work has been extended to models with more realistic synaptic kinetics (Brunel & Wang, 2003) and conductance-based models (Geisler, Brunel, & Wang, 2005).

There is also a body of work showing that stochastically driven Neural Mass Models (David & Friston, 2003) comprising stellate cells, interneurons and pyramidal cells, can generate a range of frequencies including gamma. These have recently been extended by incorporating an additional population of reciprocally connected fast interneurons (Ursino, Cona, & Zavaglia, 2010). This results in a robust model of gamma activity using realistic synaptic time constants.

For more detailed mechanisms underlying gamma oscillations we refer the reader to recent reviews (Bartos et al., 2007; Buzsaki, 2006; Wang, 2010; Whittington et al., 2000). The point here is that there is a diversity of possible network mechanisms (PING/ING, gap junctions, WSS) underlying gamma activity. We also emphasize that gamma is not a unitary phenomenon and probably has different underlying mechanisms depending on, among other factors, experimental context, computational role, anatomical location etc. Importantly, almost all of the above work has focused on sustained gamma oscillations, whereas the current paper focuses on transient gamma activity. One key distinction is that the inhomogeneities in input drive that reduce the robustness of a sustained gamma oscillation are not a problem for transient dynamics. Indeed this very feature can be made use of to signal a pattern recognition event—only when the input drives are similar will synchronization occur. This endows the Hopfield and Brody model and the WCO–TS network with the specificity necessary for pattern recognition over psychophysiological time scales.

2. Materials and methods

2.1. Electrocorticogram data

This section describes a spectral analysis applied to the ECOG data presented in Canolty et al. (2007). ECOG data was recorded from an 8-by-8 grid of electrodes placed over fronto-temporal cortex. We analyse data from a single subject, a 37 year old right handed woman with medically intractable complex partial seizures, and from a single electrode (number 49) placed over Superior Temporal Sulcus (STS).

The subject listened to three types of stimuli (i) mouth- or hand-related action verbs, (ii) acoustically matched but unintelligible nonwords and (iii) proper names which served as target stimuli. The subject was instructed to press a button using their left index finger each time they heard a proper name, but not for other stimuli.

The auditory data files ('.wav' files) were adjusted to have the same power and duration. Each nonword matched one of the words (action verbs) in duration, intensity and power spectrum, but was rendered unintelligible by removing components of the modulation power spectrum using the Modulation Transfer Function (MTF) algorithm described in Elliott and Theunissen (2009). This is based on a two-dimensional Fourier transform of the log spectrogram, after which slower time–frequency modulations are removed. This results in spectrograms which are, for example, less smooth in the time and frequency domain, as shown for example in the third row of Fig. 3. Further details of the MTF processing are given in Canolty et al. (2007). Overall, our database comprised 96 speech utterances ('.wav' files) and 96 matched nonwords.

The ECOG signals were analog filtered between 0.01 and 250 Hz, digitized at a sample rate of 2003 Hz, and high-pass filtered above 2.3 Hz to minimize heartbeat artefact. The data were then epoched from 200 ms before to 1000 ms after stimulus onset to produce $i = 1$–96 time series for words, y_{wi}, and for nonwords, y_{ni}. Each time series corresponds to processed ECOG data from a single electrode in response to a speech utterance or matched nonword.

The corresponding spectrograms $G(y_{wi}, f, t)$ and $G(y_{ni}, f, t)$ were then computed using a windowed multitaper method with window size $N = 256$ samples (0.128 s), a window offset of 32 samples (0.016 s), and time-bandwidth parameter set to $NW = 3$ (Mitra & Pesaran, 1999). The time-bandwidth parameter is the product of the number of samples N and the frequency resolution, W (in radians). Thus, $NW = 3$ produces a frequency resolution of 3.7 Hz. Each spectrogram was then log-transformed so that power changes over a wide range of frequencies would be visible on the same plot.

Fig. 1 shows the average spectrograms for words and nonwords, the average difference between them (words minus nonwords), and the significance of the difference as assessed using a two-sample t-test. These spectrograms were computed using the multitaper method described above. The figure clearly shows a burst of high-gamma activity (75–150 Hz) for words but not for nonwords. This is exactly the sort of activity that the WCO–TS model predicts should accompany recognition events (see later). The differential spectrogram

$$Y(f, t) = \frac{1}{N} \sum_{i=1}^{N} [G(y_{wi}, f, t) - G(y_{ni}, f, t)] \quad (1)$$

shown in the bottom left of Fig. 1 is the data feature we wish to explain with the WCO–TS model.

2.2. Dynamic pattern recognition model

The overall processing stream for the dynamic pattern recognition model is shown in Fig. 2 and the following subsections describe each step in the processing stream. Briefly, the steps are as follows. First, as described in Section 2.2.1, the original auditory time series (bottom row of Fig. 3) are bandpass filtered into a number of different frequency bands (3rd row of Fig. 3). Second, Occurrence Time (OT) features are extracted as described in Section 2.2.2. These correspond to the times at which power in the different bands cross specified intensity levels. Third, as described in Section 2.2.3, these OT features stimulate activity in a network of feature detectors. Each detector oscillates initially at some maximum frequency which then decreases due to spike rate adaptation, as illustrated in the second row of Fig. 3. Fourth, as described in Section 2.2.4, synaptic plasticity is assumed to have connected together word-specific ensembles that have the appropriate decay constant for each feature such that, at some point post-stimulus, the relevant features are oscillating at the same frequency. This is the 'Many-Are-Equal (MAE)' coding scheme proposed by Hopfield and Brody, and the MAE point can be seen in the left column, second row of Fig. 3. Fifth, as described in Section 2.2.5, similar firing rates cause a synchronization event in a network of Weakly Coupled Oscillators which generate a gamma burst in the LFP. This is seen for a recognized auditory word in the left column, top row of Fig. 3 but not for a nonword (right column, top row). Fig. 3 demonstrates the same concept as Fig. 2 in Hopfield and Brody (2001) but uses a WCO rather than IF network. Overall, our dynamical pattern recognition model is identical to the HB model except that the synchronization mechanism is instantiated in a WCO rather than an IF network.

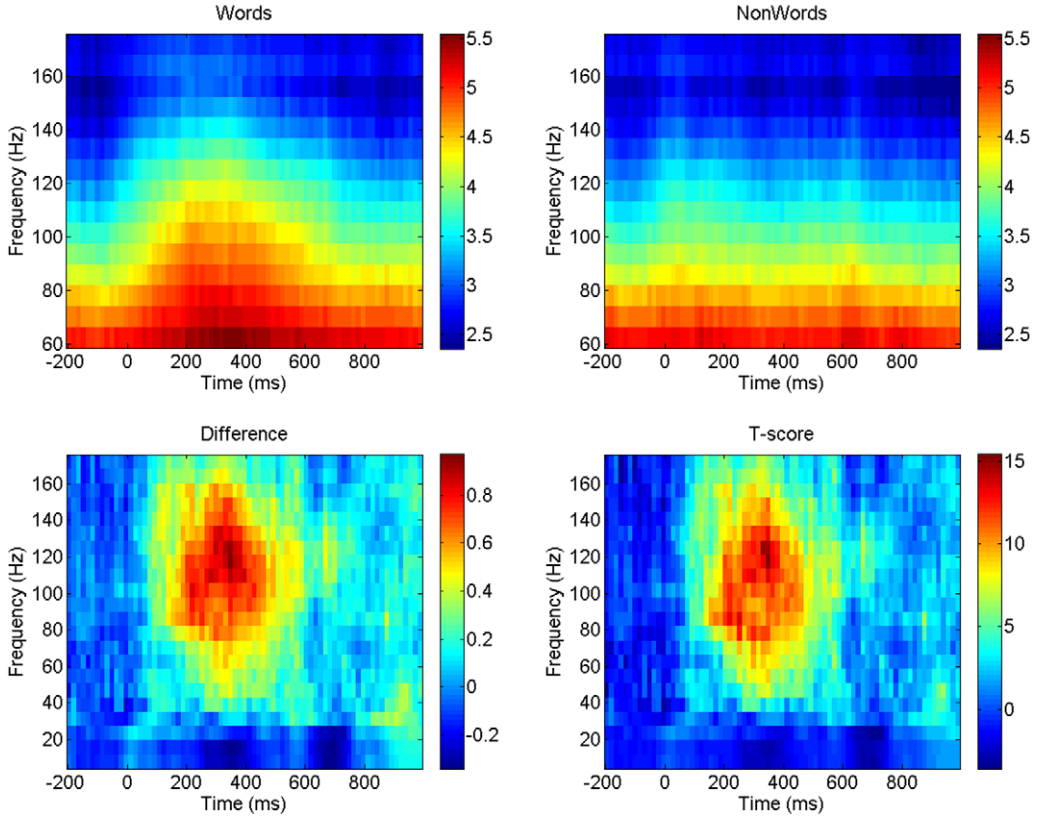

Fig. 1. (Top left) Average spectral response to words, (top right) average spectral response to nonwords, (bottom left) difference in spectral response: words–nonwords, (bottom right) significance of difference as assessed with a two-sample t-test. These spectrograms were computed using the multitaper method described in Section 2.1. *Source:* Spectral Analysis of ECOG data from Canolty et al. (2007).

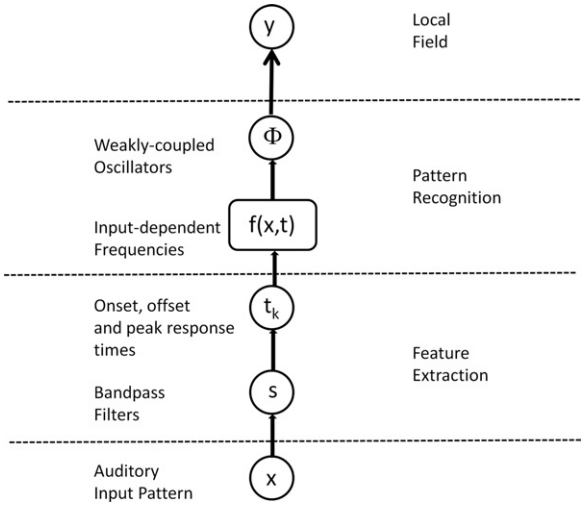

Fig. 2. The figure shows the overall processing stream for the dynamic pattern recognition model. This comprises feature extraction, pattern recognition and forward modelling of neuroimaging data. The weakly coupled oscillator dynamics are implemented using Eq. (3). These are driven by stimulus-dependent input frequencies described using Eqs. (2) and (7).

2.2.1. Bandpass filters

Our model assumes that neural circuits in a feature detection region are tuned to specific frequency ranges in a manner that is broadly similar to processing in the cochlear and basilar membrane of the mammalian auditory system. We characterize this activity with $b = 1..B$ frequency bins between $f_{\min}$ and $f_{\max}$ Hz where ω_b is the centre frequency of the bth bin. Power in each is computed by bandpassing the input time series, $x(t)$, and then computing the Hilbert envelope $s_b(t)$. Bandpass filters were implemented using Finite Impulse Response (FIR) filters of order 80. Filter coefficients were computed using a least squares criterion and the filters were applied in forward and reverse directions to obtain zero-phase distortion. The filtering was implemented using the `firls.m` and `filtfilt.m` functions from the Matlab signal processing toolbox. We emphasize here that these filters are applied to the auditory stimuli rather than to the ECOG data. Example bandpass filter responses to auditory input are shown in the third row of Fig. 3. More physiologically realistic filters can be implemented by linear spacing the filter bands on a mel-frequency scale (Ghitza, 1986), and this was implemented for the pattern recognition results described in Section 3.1.

2.2.2. Occurrence Times

Neurons or neural circuits then respond to three types of features within each frequency band: onsets, offsets and peaks of activity. Such frequency tuned onset and offset detectors have been observed in the inferior colliculus of the auditory midbrain (Casseday et al., 2002). Onset and offset times are computed from the first and last crossings of $s_b(t)$ with a fixed threshold. The peak time is computed from the maximum value of $s_b(t)$. Overall, in response to input pattern x we have K features which are detected at times $t_k(x)$ with $k = 1..K$. For the analysis of the ECOG data (see below) we only use those features for which t_k is less than 300 ms, as recognition is required before the end of the word.

Greater physiological realism can be added by using multiple level crossings to define multiple onset and offset points, as in (Ghitza, 1986; Gutig & Sompolinsky, 2009). This adds the property of intensity coding of the auditory signal and was implemented for the pattern recognition results described in Section 3.1. A similar encoding scheme has been proposed by Loiselle, Rouat, Pressnitzer, and Thorpe (2005).

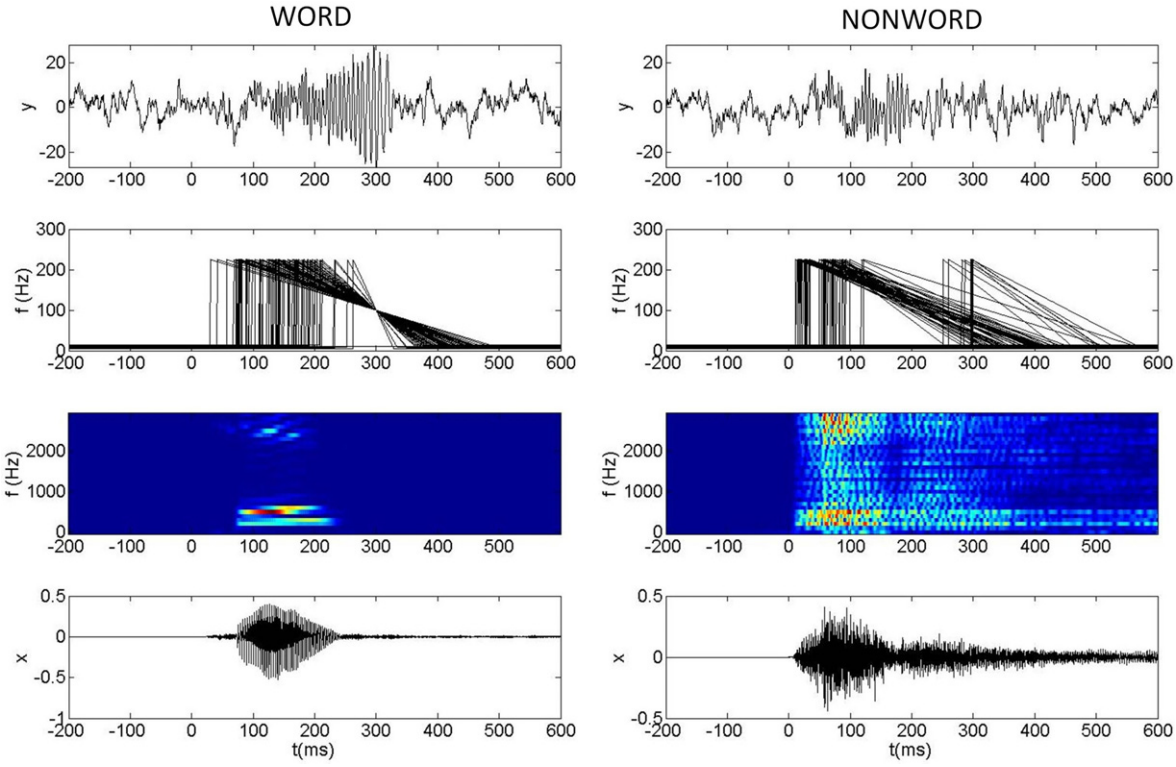

Fig. 3. The figure shows auditory inputs $x(t)$ (bottom row), auditory spectrograms $s(t)$ (third row), input frequencies to the WCO–TS network $f_k(x, t)$ (second row), and LFP signals from the WCO–TS net (top row) for a word (left column) and nonword (right column). The times at which the frequencies ramp up to their maximal value (in the second row) are the Occurrence Times (OTs). The word is 'Hiss' and the nonword was produced using MTF filtering (see text). The adaptation constants τ_j have been optimized such that the input frequencies become equal at $t_b = 300$ ms for the word input. The same τ_j's are used to generate the input frequencies for the nonword. For the nonword input there is no time point at which all the frequencies are equal and consequently no large LFP gamma burst.

2.2.3. Spike rate adaptation

When a feature is detected the relevant neuron or neural circuit responds by firing at a high frequency, $f_{\max}$, which then decreases. At the single cell level this is known as Spike Rate Adaptation (SRA) or Spike Frequency Adaptation (SFA). Timescales of decay range from tens of milliseconds in the auditory nerve (Zhang, Miller, Robinson, Abbas, & Hu, 2007) to several seconds for delay activity in frontal cortex. Optical imaging reveals larger time windows of temporal integration as one moves from primary to secondary auditory areas (Harrison, Harel, Panesar, Mori, & Mount, 2000). In primary auditory cortex, Ulanovsky, Las, Farkas, and Nelken (2004) have observed within-trial adaptation time constants, at a fast 10 ms time scale, and a slower 150 ms scale. It is these longer time constants that are hypothesized to be useful for auditory object recognition (May & Tiitinen, 2007).

Following HB, for each feature detector we envisage $j = 1..J$ neurons or neural circuits each with a different decay constant, τ_j. For the jth circuit detecting the kth feature we assume these frequencies are given by linear decays

$$f_{kj}(x, t) = f_{\max}\left(1 - \tau_j[t - t_k(x)]\right) h[t - t_k(x)] \quad (2)$$

where $h[a]$ is the threshold function with $h[a] = 1$ for $a \geq 0$ and zero otherwise. The maximum firing frequency is $f_{\max}$. We have also experimented with exponential decay functions but found that linear functions produced gamma bursts that are better localized in time (and better match ECOG data—see Section 3.2). Overall, the feature detection region comprises $C = K \times J$ feature detectors.

2.2.4. Synaptic plasticity

If the input pattern, x, is a word we assume that synaptic plasticity will have acted so as to connect K out of C oscillators together with uniform coupling strength $w_{kk'} = A$. This ensemble will be specialized for recognizing a particular word. Different words will then activate different ensembles of size K in auditory cortex. This is broadly consistent with electrophysiological recordings from non-human primates where representations are composed of small dynamic subsets of highly active neurons (Hromadka, Deweese, & Zador, 2008).

We are effectively assuming, following HB, that for each feature there are $j = 1..J$ neurons or neural circuits that respond initially at high frequency $f_{\max}$, and then with linearly reducing frequency specified by decay constants τ_j. The role of synaptic plasticity is to choose the optimal τ_j for each feature so that there will be a poststimulus timepoint at which the frequencies become equal. This concept is described in Abbott (2001), and illustrated in Fig. 4. For example, if a word contains early onset of 1 kHz activity, then a long τ_j will be selected for that feature. If it contains late offset 1 kHz activity then a short τ_j will be selected for that feature.

In this paper we implement the plasticity process by simply computing those values of τ_j that will, given an initial frequency $f_{\max}$, make all the K frequencies equal to f_b at time t_b. We refer to these optimal values as $\tau_{j_{\text{opt}}}$. In the HB model it is spike timing that is synchronized and it is proposed that $\tau_{j_{\text{opt}}}$ can be learnt via spike timing dependent plasticity (Lee, Sen, & Kopell, 2009). More recently, Gutig and Sompolinsky (2009) have shown that this can be implemented using a conductance-based tempotron. They also provide an analysis of the storage capacity of the HB-type coding scheme, estimating that it can store 0.0625 patterns per synapse. Thus, to store 5000 words would require 80,000 synapses.

2.2.5. Time-warp invariance

Human speech is characterized by a four-fold variation in the speed at which words are spoken (Miller, Grosjean, & Lomanto, 1984), and any speech recognition system whether artificial or

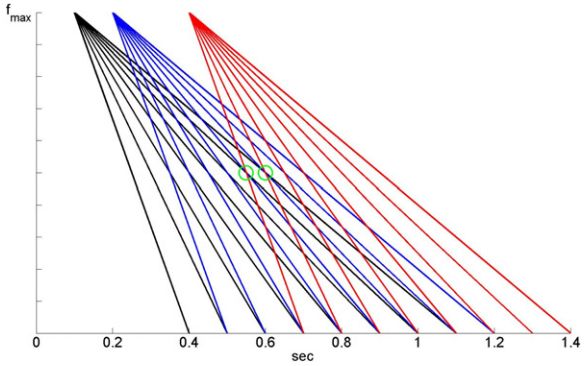

Fig. 4. The figure shows the detection of three different input features, $k = 1, 2, 3$ coloured black, red and blue. The kth feature is detected at time t_k—these are the Occurrence Times (OTs). For each feature there are $j = 1..J$ cells or circuits that initially respond at frequency $f_{\max}$, and then with linearly reducing frequency. The slopes of the frequency reduction are specified by the constants τ_j. The role of synaptic plasticity is to choose the optimal τ_j for each feature such that there will be a poststimulus timepoint, t_b, at which the frequencies become equal (f_b). These points are indicated by the green circles. Generally, plasticity acts to select long decay constants for early features and short ones for later features. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

natural, will have to deal with this range of 'time-warp'. In the above coding scheme time-warp invariance is achieved because the timing of the recognition event (gamma burst) depends on the speed at which the word is spoken. This is discussed at length in Hopfield and Brody (2001) and illustrated in Fig. 5. Time-warp invariance occurs rather naturally with OT features and WCO–TS/HB models but is more complicated to add to other representations. ASR based on cepstral coefficients, for example, requires an additional Dynamic Time Warping or Hidden Markov Modelling stage (Rabiner & Juang, 1993). What we have described so far is identical to the HB model. In fact, our dynamical pattern recognition model is the same as HB, except that the synchronization mechanism is instantiated in a WCO rather than an IF network, as described below.

2.2.6. Weakly coupled oscillator network

Weakly coupled oscillators are a standard approach for studying synchronization dynamics (Hoppensteadt & Izhikevich, 1997). They can be derived by applying a phase reduction approach to neurophysiologically realistic neural network 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 et al., 1995). Cortical neuron models, for example, such as the Quadratic Integrate and Fire model or any model with 'type 1' dynamics (Hoppensteadt & Izhikevich, 1997), can be implemented as a 'theta neuron'. This is a differential equation with a single phase variable, and a particular 'phase interaction function'. Networks of such cells can then be analysed to find the neuronal mechanisms that give rise to synchronization. An early application, for example, used WCOs to infer that neuronal inhibition rather than excitation can cause synchronous activity (Vreeswijk et al., 1994).

More abstract models based on WCOs have also been used as neurocognitive models of visual attention (Corchs & Deco, 2001) and attention-guided object selection (Borisyuk & Kazanovich, 2004). These were based on previous models of visual attention (Niebur & Koch, 1994) and image segmentation (Wang & Terman, 1997) that also made use of synchronizing dynamics.

In the HB model it is the action potentials of IF neurons that become synchronized. Synchronization is brought about by balanced excitation and inhibition among excitatory and inhibitory IF cells. In this paper we are equivocal about the details of local neurons or neural circuits that bring about oscillation and synchronization (for reasons discussed above). Instead, we consider the properties of cells or circuits of cells using a description at the level of phase dynamics. Other than this difference, our model is more or less identical to that in HB.

It is assumed, as it is in the HB model, that Spike Timing Dependent Plasticity (STDP) is the underlying mechanism for forming ensembles of cells that synchronize together. Because cells must spike within a maximal period of 25 ms (approx) then the minimum frequency of the network oscillation is 40 Hz. Tighter synchronization will lead to higher frequency rhythms. This assumes that ECOG is detecting LFP activity of synchronized onset/offset detectors.

In a network of $k = 1$ to K oscillators, each oscillates at unit amplitude with frequency $f_k(x, t)$ and phase $\phi_k(x, t)$ where x denotes the stimulus pattern and t denotes time. The rate of change of phase of the kth oscillator is given by

$$\dot{\phi}_k(x, t) = f_k(x, t) - \sum_{k'=1}^{K} w_{kk'} h[\phi_{k'}(t) - \phi_k(t)] + e_k(t) \tag{3}$$

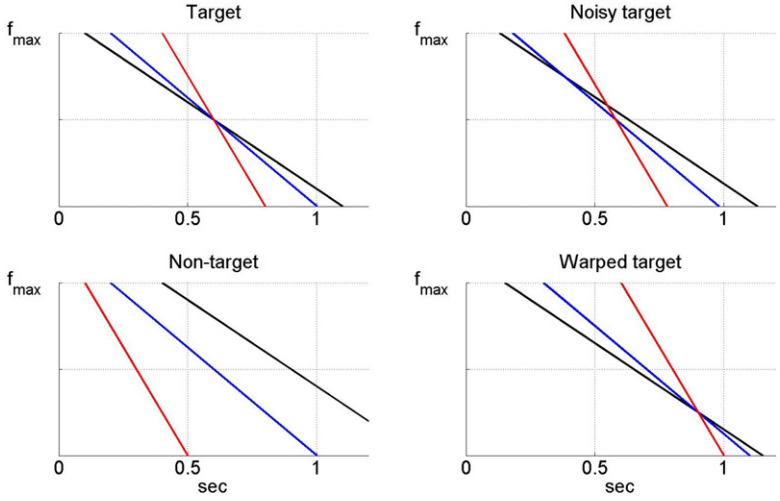

Fig. 5. The top left figure shows the oscillator frequencies selected by the rightmost green circle in Fig. 4. The top right shows the same responses but for 5% noise added onto Occurrence Times. The bottom left shows the responses for a non-target pattern. Here, the 'red' feature now occurs before the 'black' feature so that there are no timepoints when the frequencies are similar. The bottom right figure shows responses for the target pattern but where Occurrence Times have been multiplied by 1.5. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

where $w_{kk'}$ is the coupling strength between oscillators k and k', $h[\Delta\phi]$ is a phase interaction function, and $e_k(t)$ is additive circular Gaussian noise and $f_k(x, t)$ is the frequency of the kth oscillator. In this paper the frequencies $f_k(x, t)$ are nonstationary and depend on the stimulus pattern x.

In this paper we make the simplifying assumption that the Phase Interaction Function (PIF) $h(\Delta\phi) = \sin(\Delta\phi)$ which results in the Global Zero Lag (GZL) solution (where all phase differences are zero—i.e., full synchronization), being a potential stable state of the system (Ermentrout & Kleinfeld, 2001; Penny et al., 2009). We envisage that in future work it might be possible to infer h based on neuroimaging data, an approach we have implemented for magnetoencephalograph data (Penny et al., 2009).

The noise is drawn from a circular Gaussian (von-Mises) density with zero mean and precision κ. This particular form was chosen as it is the simplest density that gives rise to noise that is bounded between 0 and 2π. The inclusion of noise was found to provide better fits to the ECOG spectrograms (see Section 3.2). For the results below we used a value of $\kappa = 10$ and used 'frozen noise' by drawing $e_k(t)$ prior to each model fitting run.

WCO models are often studied with the assumption that the frequencies are stationary, $f_k(x, t) = f_k(x)$. Much research examines the robustness of stable synchronized states as a function of variations in oscillator frequencies. Kuramoto (1984), for example, has derived the following result. If the frequencies are Lorentz distributed with variance σ_L^2, then a stable synchronized state will be reached if the coupling parameters $w_{kk'} = w$ and satisfy

$$\frac{\sigma_L}{wK} < \frac{1}{2}. \tag{4}$$

This shows that if the frequencies are more heterogeneous then stronger coupling is needed to reach a synchronized state. It is not specified, however, how long it takes for synchronization to be achieved so this result is not of immediate relevance to the WCO–TS model.

In previous work we have used WCOs to study synchronization among different brain regions (Penny et al., 2009) whereas in this paper we use them to model activity in a single region. Specifically, the Local Field Potential (LFP) in a region is modelled as

$$y(x, t) = \sum_{k=1}^{K} \cos[\phi_k(x, t)]. \tag{5}$$

If the oscillators are phase aligned then the field activity will reach a maximum value K. Weaker synchronization results in a smaller LFP. This is shown in the top row of Fig. 3 where we have a large gamma burst in response to a word and a small one in response to a nonword (a pattern which the network has not previously been exposed to).

A quantitative relationship can be derived, for example, by assuming that the instantaneous phases are Gaussian distributed with phase variance σ_t^2. The instantaneous field power is then given by (Roweis, 2009)

$$\langle y(x, t)^2 \rangle = 1 + e^{-\sigma_t^2}(K - 1). \tag{6}$$

This completes the description of the WCO network.

2.3. Modelling electrocorticogram data

For modelling the ECOG data, the input frequency to the kth oscillator in the WCO network is given by

$$f_k(x, t) = f_{kj_{\text{opt}}}(x, t) + g_k \tag{7}$$

where $f_{kj_{\text{opt}}}(x, t)$ is defined in Eq. (2) with $j = j_{\text{opt}}$, and g_k the baseline frequency of the kth oscillator. This baseline determines the oscillation frequency before and after the stimulus-induced transient. We define an average baseline frequency for the ensemble, g, and then draw g_k from a uniform distribution centred on g and with a bandwidth of $g/2$. In this paper we use $g = 10$ Hz so as to reflect typical background alpha activity (Buzsaki, 2006).

Critically, if the input pattern x is a nonword then we assume that the τ_j's will not have been optimized for the features $t_k(x)$. For nonwords we use the τ_j's associated with the corresponding paired word. For nonwords the input frequencies $f_k(x, t)$ will then not be all equal at time t_b and there will be no LFP gamma burst (or it will be greatly reduced in power). This is illustrated in the top two rows of Fig. 3.

To fit the ECOG spectrogram we wish to obtain the difference in spectral responses between words and nonwords. This could be implemented by computing the field activity for all C oscillators in the pattern recognition region. However, in this paper we make a computational saving by considering only the field contribution from those cells that are activated by both stimulus patterns of a given word/nonword pair (note that for a given input not all cells may be activated as the onset/offset/peak response may occur after the cut-off time of $t_k \leq 300$ ms—see above)

In early experiments with the WCO–TS model we became concerned that bursts of activity would also emerge at resting frequencies. We therefore considered an augmented model in which coupling parameters were allowed to be frequency dependent, as described in the Appendix B (but see results).

2.3.1. Spectrograms

We choose $i = 1$ to N pairs of word exemplars, x_{wi}, and nonword exemplars, x_{ni}. For each pair we compute the local field responses $y(x_{wi}, t)$ and $y(x_{ni}, t)$ by first integrating the WCO dynamics (Eq. (3)) using the Euler–Maruyama method (Kloeden & Platen, 1999) to compute the phase time series for the network of oscillators, ϕ_k, and then use Eq. (5) to produce the field response. Examples of these LFPs are shown in the top row of Fig. 3.

We also considered an augmented model in which burst time and frequency (t_b and f_b) were allowed to vary over trials. We considered variations of the form

$$t_b^i = (1 - \delta t)t_b + 2\delta t z_i \tag{8}$$
$$f_b^i = (1 - \delta f)f_b + 2\delta f w_i$$

where δt and δf are parameters to be estimated, and $\{z_i, w_i\}$ are random variables uniformly distributed between 0 and 1.

The corresponding spectrograms $G(x_{wi}, f, t)$ and $G(x_{ni}, f, t)$ are computed using the windowed multi-taper method described above, using identical parameters as for the ECOG data itself. It is then possible to compute the average spectral difference between word and nonword responses

$$D_s(f, t; \theta) = \frac{1}{N}\sum_{i=1}^{N}[G(x_{wi}, f, t) - G(x_{ni}, f, t)] \tag{9}$$

where θ are model parameters (see below). Upon analysing the auditory data files we noticed a systematic bias in mean OTs between words and nonwords (123 ms for nonwords versus 140 ms for words, $p = 0.04$). As we did not wish this to unduly influence the synchronization processes we adopted the following procedure. We first define the spectral difference that would be obtained without any synchronization (this is obtained by setting the coupling $A = 0$)

$$D_0(f, t; \theta) = \frac{1}{N}\sum_{i=1}^{N}[G_0(x_{wi}, f, t) - G_0(x_{ni}, f, t)]. \tag{10}$$

The predicted spectral difference from the WCO–TS model is then given by

$$D(f, t; \theta) = D_s(f, t; \theta) - D_0(f, t; \theta). \tag{11}$$

Table 1
Minimal model uniform priors over model parameters. The minimal model used four parameters only: burst time (t_b), burst frequency (f_b), maximum frequency ($f_{\max}$), and coupling strength (A).

Parameter	Minimum	Maximum
t_b (ms)	200	300
f_b (Hz)	120	150
$f_{\max}$ (Hz)	150	250
A	−3	0

Table 2
Augmented model uniform priors over model parameters. Augmented models used the same parameters as for the minimal model but additionally had parameters for (i) frequency dependent synchronization (β_1, β_2) and/or (ii) between-trial burst time/frequency variability (δt, δf). The δ_t and δ_f parameters, for example, allow for between a 10% and 30% trial to trial variability in burst time/frequency.

Parameter	Minimum	Maximum
β_1	−0.25	0.25
β_2	0.6	0.9
δt	0.1	0.3
δf	0.1	0.3

2.3.2. Parameter estimation

Our model contains stochastic variables such as the state noise, $e_k(t)$, and for the augmented models, the between trial variables $\{z_i, w_i\}$. These random variables were drawn prior to each model fitting run. The use of such 'frozen noise' ensures that models with the same parameters have the same likelihood, a requirement of the model fitting procedure.

The other parameters, θ, (see Tables 1 and 2) were estimated from data. This data comprise a subset of the (paired) word and nonword utterances (auditory inputs and ECOG spectrograms). The overall set of 91 exemplars was split into a training set of 42 exemplars, used to estimate model parameters, and a test set of 51 exemplars (the other 3 outlying exemplars were removed). Model fitting was implemented using a Bayesian parameter estimation algorithm as follows.

To obtain a quantitative measure of how well the data and model spectrograms are matched we first normalize both to have unit power over the specified time interval. We then project $D(f, t)$ onto the first eigenvector of $Y(f, t)$. Model error, $E(\theta)$, is then computed as the Root Mean Squared (RMS) difference between the resulting projections. Log model likelihood is defined to be proportional to negative model error

$$p(Y|\theta) \propto \exp\left[-E(\theta)\right]. \tag{12}$$

We placed uniform priors over model parameters as shown in Tables 1 and 2. The posterior density was then estimated using a Metropolis–Hastings (MH) algorithm (Gelman, Carlin, Stern, & Rubin, 1995). This uses a proposal density which we chose to be a zero-mean diagonal covariance Gaussian with standard deviation for the pth parameter given by

$$\sigma_p = \frac{\theta^p_{\max} - \theta^p_{\min}}{S} \tag{13}$$

where $\theta^p_{\min}$, $\theta^p_{\max}$ are defined in Tables 1 and 2, and S was chosen to achieve high acceptance rates (see results). At each iteration of the MH algorithm a proposal, θ^*, is generated by adding a sample from the proposal density onto the sample from the previous iteration θ^{n-1}. Proposals that fell outside the range of the uniform prior were immediately rejected. Other proposals were accepted with probability

$$\min\left(\frac{p(\theta^*|Y)}{p(\theta^{n-1}|Y)}, 1\right). \tag{14}$$

If the proposal is accepted we set $\theta_n = \theta^*$, and if not then $\theta_n = \theta_{n-1}$. The above MH criterion ensures that, after a burn in period, the algorithm produces samples from the posterior density of interest (Gelman et al., 1995). This posterior distribution (as with any Bayesian estimation) takes into account both the uniform priors over model parameters (see Tables 1 and 2) and the fit of the model to the data.

3. Results

The first results section examines the suitability of OT features for speech recognition, by using the multiple speaker isolated word database described in Hopfield and Brody (2000, 2001). The second results section examines the use of the overall dynamic pattern recognition process as a forward model of Electrocorticogram data.

3.1. Pattern recognition

We now demonstrate the usefulness of Occurrence Times (OTs) as speech recognition features by comparing them to a more commonly used feature, Mel-Frequency Cepstral Coefficients (MFCCs) (Roweis, 1998). MFCCs form the front-end of state-of- the-art speech recognition systems such as HTK (Woodland, 2003) and Sphinx-4 (Walker, Lamere, Kwok, Raj, & Singh, 2004). To simplify the comparison of MFCC and OT features they were both fed into an identical pattern recognition stage, chosen to be a first nearest neighbour classifier (Duda & Hart, 1973).

We used the multiple speaker isolated word database described in Hopfield and Brody (2000, 2001). This comprises 500 speech files, the words 'zero', 'one', through to 'nine', spoken by five different female speakers with ten replications of each word per speaker. This data is a subset of the TI46 database from the Linguistic Data Consortium (Verstraeten, Schrauwen, Stroobandt, & Campenhout, 2005). The data was partitioned into a fixed training set, comprising 5 utterances per digit per speaker, and a fixed test set holding the remaining 5 utterances per digit per speaker. This gives 250 training and 250 testing exemplars. We first report recognition results on this noiseless isolated word data set and then go on to test the systems in the presence of additive noise. The Word Error Rates (WERs) reported below refer to the error rate on the test set.

The MFCCs were computed using a standard processing pipeline as described in the Appendix A. One issue here arises from the fact that the speech time series must first be partitioned into frames, and the overall MFCC vector is then concatenated over frames. As the speech signals were, however, of different lengths and comparison of feature vectors (see below) requires them to be of the same length we derived features for a fixed time period t_{fix} ms. Blocks for which signals were unavailable (e.g., due to words being shorter than t_{fix} ms) were assigned MFCC values of zero. This coding also provides discriminatory information (e.g., that word A is shorter than word B). We also tried setting t_{fix} to the length of the shortest word in the database, but this resulted in much worse classification performance. We varied various parameters of the MFCC processing to achieve optimal performance. The best performance, a Word Error Rate (*WER*) of 0.000 (perfect discrimination), was achieved with $K = 18$ cepstral coefficients per frame. Other parameters were set as described in the Appendix A.

The fact that different speech signals were of different length provided no problem for OT features as a fixed length code was always readily obtainable (each channel always has an onset, peak and offset regardless of the amount of data). The onsets, peaks and offsets were defined as described earlier. We also implemented robustness to global time-warps by dividing all OT values by the time between the first and last OT.

We used a fixed number $B = 11$ frequency bins. This system produced $WER = 0.100$. We then improved the system by introducing multiple level detectors for each onset and offset feature,

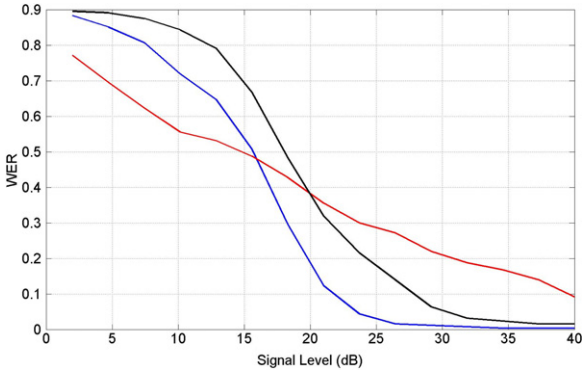

Fig. 6. Speech recognition in additive white noise. We plot Word Error Rate (WER) against signal level for optimized speech recognition systems using Occurrence Time (OT) features (red curve), Mel Frequency Cepstral Coefficients (MFCC) (blue curve) and MFCC coefficients but with the number of features matched to that of the OT system (black curve). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

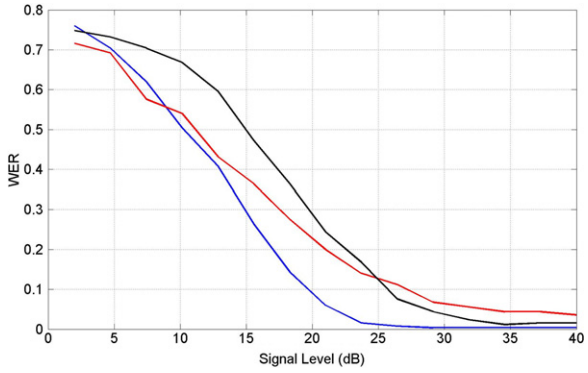

Fig. 7. Speech recognition in additive speech babble. We plot Word Error Rate (WER) against signal level for optimized speech recognition systems using Occurrence Time (OT) features (red curve), Mel Frequency Cepstral Coefficients (MFCC) (blue curve) and MFCC coefficients but with the number of features matched to that of the OT system (black curve). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

as proposed in Gutig and Sompolinsky (2009). This additional 'intensity encoding' reached an optimal level of performance with 7 intensity levels per detector, with $WER = 0.024$. Thus, for each of the 11 frequency bands there is one maximum detector, 7 level crossings for onsets and 7 level crossings for offsets. Overall, that is 15 features per frequency band giving a total of 165 features. In the original HB paper it was proposed that performance might be further improved if multiple crossings of each intensity level were allowed. We implemented this but did not find a reduction in WER on this database.

Thus, on the noiseless data we can conclude that OT features with multiple levels of intensity encoding provide reasonable recognition performance, though not as good as MFCCs. We now take the optimized MFCC and OT systems and apply them to noisy data. We first corrupted the test data with white noise to produce a range of signal to noise ratios. We define the signal level, measured in decibels (dB), as $S = 20\log_{10}(\sigma_s/\sigma_e)$ where σ_s is the signal standard deviation and σ_e is the noise standard deviation. Fig. 6 shows that speech recognition performance rapidly degrades at less than 25 dB, as is well known (Ghitza, 1986). It also shows that the MFCC system is better for high signal levels whereas the OT system is better for low signal levels.

We thought it might be possible that the OT system performed better at low signal levels because it had fewer parameters than the MFCC system, and so might generalize better (Bishop, 1995). We therefore applied an MFCC system that was matched in the number of parameters, but this did not improve performance at low signal levels (see black curve in Fig. 6).

Finally, we repeated the speech recognition tests with additive speech 'babble' which was derived from 100 people speaking in a canteen (this data was downloaded from the Signal Processing Information Base http://spib.rice.edu). The results in Fig. 7 show that MFCC is better for high signal levels but that OT and MFCC have the same performance at low signal levels.

We conclude that OT features provide a compact code for auditory word recognition that rivals that of standard encoding methods in noisy environments. An important caveat here is that we have used the same back-end for both OT and MFCC features, namely a nearest neighbour classifier. This back-end is not optimal for either front-ends, but serves to provide a common baseline for both approaches. MFCC features are much more powerful when used in combination with an HMM classifier (Woodland, 2003), and OT features when used with a recognition process such as a Tempotron (Gutig & Sompolinsky, 2009). We return to this issue in the discussion.

3.2. Modelling electrocorticogram data

This section describes the use of the WCO–TS network as a forward model of ECOG data. The networks are driven by auditory data (word versus nonword). As described in Section 2.1, this data set comprises $i = 1..96$ word utterances and $i = 1..96$ paired nonword utterances. For each utterance we have the original auditory data file (y_{wi} and y_{ni} for words and nonwords) and a spectrogram of the corresponding ECOG response, $G(y_{wi}, f, t)$ and $G(y_{ni}, f, t)$.

For each utterance, the auditory signal, $x(t)$, produces band-pass filtered responses (described in Section 2.2.1), which in turn produce Occurrence Time features (Section 2.2.2) in a bank of feature detectors. Each detector oscillates initially at some maximum frequency which then decreases due to spike rate adaptation. Oscillator synchronization then produces a field potential from the WCO–TS network as described in the remainder of Section 2.2. The data set of (paired) word and nonword utterances (96 exemplars, comprising auditory inputs and ECOG spectrograms) are then split into a training set of 42 exemplars, and a test set of 51 exemplars (3 outlying exemplars were removed).

The Bayesian parameter estimation algorithm (Section 2.3.2) was run on data from the 42 training exemplars. For the proposal densities in the Metropolis–Hastings algorithm we used a value of $S = 8$ as this achieved a high acceptance rate. We ran the algorithm for 2000 iterations and discarded samples from the first 1000 iterations. For each sample, approximately 15 s of computer time were required to compute the likelihood (using a 64-bit dual core IBM with 12 G RAM and 3.2 GHz clock speed). Overall, model fitting required about 8 h of computer time. Model fits were assessed using a likelihood function that is related to the RMS error between training data and model spectrograms (see Eq. (12)).

A numerical benefit of describing transient synchronization by phase dynamics (Eq. (3)), rather than more detailed IF network models (Hopfield & Brody, 2001), is that they describe a similar phenomenon but are quicker to numerically integrate. For the WCO integration we used a step size of 2.5 ms (the limit on the step size is due to the sampling rate we need to estimate the frequencies produced).

We first present results from the minimal model (model M0) which used just four adaptable parameters ($t_b, f_b, f_{\max}, A$). Fig. 8 shows posterior distributions of these parameters as estimated using the MH algorithm. The fairly flat distributions indicate that data fit is relatively insensitive to the exact values of these parameters. This is a sign of a good model. Fig. 9 (left column) shows the data and model spectrograms for the maximum posterior sample.

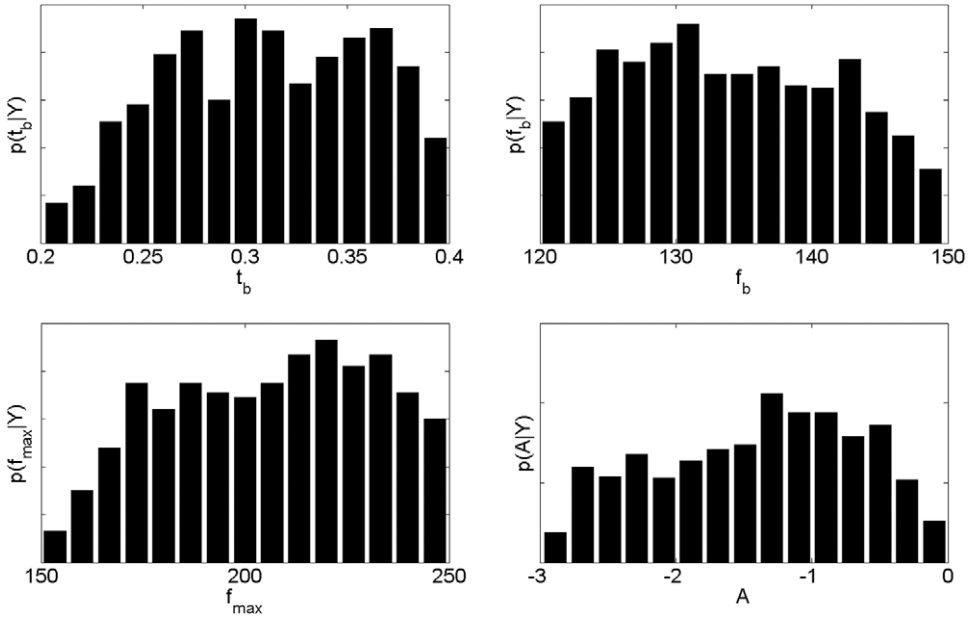

Fig. 8. Posterior densities over burst time t_b, burst frequency f_b, coupling parameter A, and maximum frequency $f_{\max}$. The values in the histograms sum to unity as they depict probability densities.

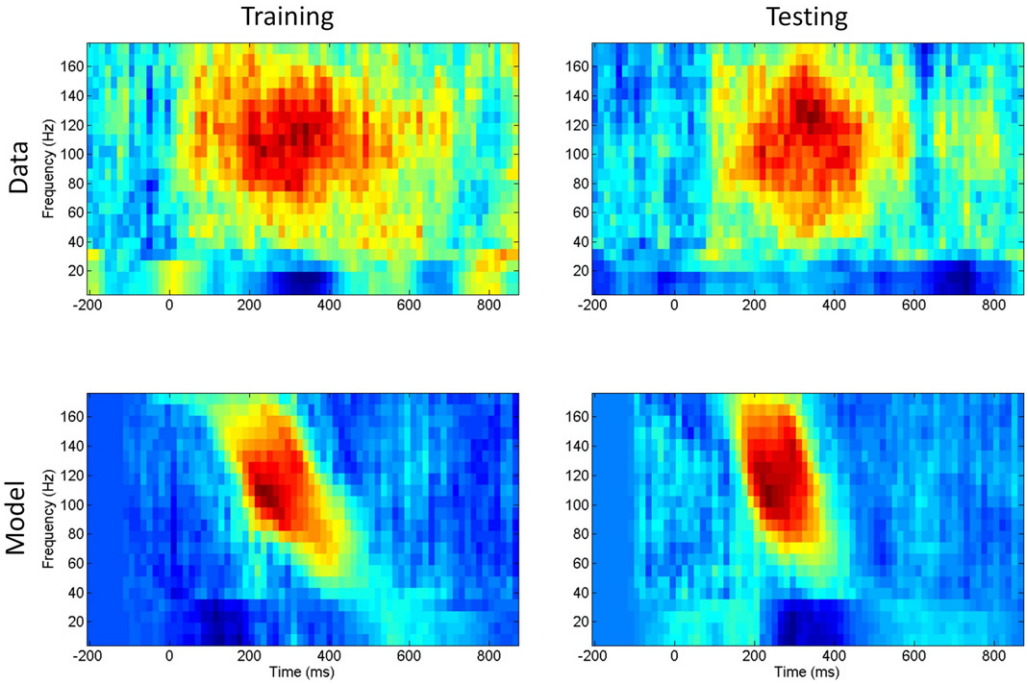

Fig. 9. *Minimal model* Top row: data spectrogram for training set (left) and test set (right). Bottom row: model spectrogram for training set (left) and test set (right) computed using a high-likelihood sample from the posterior density.

The performance of the model was then evaluated on an independent test set comprising 51 exemplars (see above). Data and model spectrograms for the test examples are shown in Fig. 9 (right column). Both training and test data show good agreement between data and model spectrograms.

We now present the results of an augmented model (model M2) which also contained frequency dependent synchronization and between-trial burst time/frequency variability (see Table 2). Fig. 10 shows the data and model spectrograms.

We also compare augmented and minimal models using the model evidence, as computed using the Posterior Harmonic Mean (PHM) (Gelman et al., 1995). This approximates the evidence for a model using samples from the posterior density

$$p_{\mathrm{PHM}}(Y|M) = \left[\frac{1}{N_s}\sum_{n=1}^{N_s}\frac{1}{p(Y|\theta_n, M)}\right]^{-1} \tag{15}$$

where $p(Y|\theta_n, M)$ is the likelihood of the nth posterior sample, and $N_s = 1000$. We compared the minimal model (M0) to two augmented models, one with between-trial burst time/frequency variability (M1), and one with both frequency dependent synchronization and between-trial burst time/frequency variability (M2). The resulting Bayes factors were 0.98 and 0.99 indicating that neither of the augmented models are significantly better than the

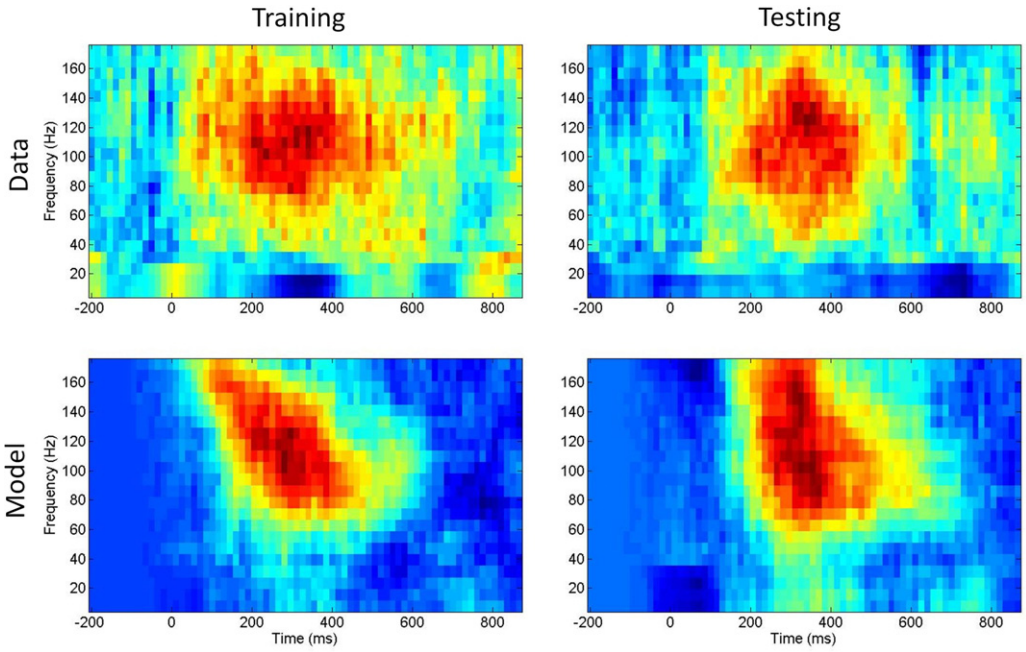

Fig. 10. *Augmented model* Top row: data spectrogram for training set (left) and test set (right). Bottom row: model spectrogram for training set (left) and test set (right) computed using a high-likelihood sample from the posterior density.

minimal model (this would require a Bayes factor of at least three (Gelman et al., 1995)).

4. Discussion

This paper has described 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 work is based heavily on prior developments by Hopfield and Brody (2001) and we have made three novel contributions to the literature.

First, we have viewed the HB model at two separate levels of analysis; the algorithmic and implementation levels (Marr & Poggio, 1976). Algorithmically, the HB model is marked out by its use of Occurrence Times (OTs) as features. We have shown using a nearest neighbour classifier that, for noisy recognition environments, OT features rival standard MFCC features in classification accuracy. For non-noisy data MFCC features were found to be better.

An important caveat to the above finding is that we used the same back-end for both OT and MFCC features, namely a nearest neighbour classifier. Moreover, this back-end is not optimal for either front-ends, but serves to provide a common baseline for both approaches. MFCC features are much more powerful when used in combination with an HMM classifier (Woodland, 2003).

For example, an MFCC–HMM system can achieve a word error rate of only 11% for connected digits in noise, with an additive noise level of 10 dB (Lee, Glass, & Ghitza, 2011). A similar level of performance is obtained when using ecologically realistic noise samples from the Aurora-2 database (Pearce & Hirsch, 2000). This is to be contrasted with the relatively poor performance of the MFCC–NN system obtained in this paper on the simpler recognition problem of isolated digits in noise (see Figs. 6 and 7), where we obtain a word error rate of about 70%. We should also bear in mind, however, that our results were obtained by training the system on clean utterances whereas the results in Lee et al. (2011) were obtained from a system trained on noisy utterances. Our results are more in line with those of Rouat, Loiselle, and Molotchnikoff (2011) who obtained word error rates of 78% when training an MFCC–HMM system on clean utterances and testing it on 10 dB noisy utterances using noise samples from Aurora-2.

We also note that OT features are also more powerful when used with a matched, optimized recognition process such as a Tempotron. An OT-Tempotron (Gutig & Sompolinsky, 2009) matched the performance level of MFCC–HMM approaches implemented in the HTK (Woodland, 2003) and Sphinx 4 (Walker et al., 2004) ASR systems, on noiseless isolated word recognition.

Second, we have proposed a generic model of transient synchronization based on Weakly Coupled Oscillators. This has allowed us to focus on the dynamics of transient synchronization per se rather than on the neural mechanisms by which the underlying oscillations are generated. As the way in which one cell or circuit couples with another can be summarized using 'phase interaction functions' (Penny et al., 2009) we envisage that it should be possible to identify families of neurons or neural circuits that have the appropriate synchronization properties.

Third, we have shown that the dynamical pattern recognition process can act as a forward model of neuroimaging data. Previous studies in this area (Corchs & Deco, 2004; Husain, Tagamets, Fromm, Braun, & Horwitz, 2004) have used computational models of auditory processing as forward models for fMRI data. Corchs and Deco (2004), for example, have used a neurodynamical model of feature-based attention in combination with a haemodynamic process as a forward model of fMRI activity. Husain et al. (2004) have taken a similar approach using a large-scale neural network of the auditory system as a model of fMRI activity. We have used ECOG data which, having a higher temporal resolution than fMRI data, has allowed us to focus on the dynamics on the recognition process itself. We were also able to show how the parameters of our model can be directly fitted to neuroimaging data using Bayesian inference. This follows the example set by Dynamic Causal Modelling (Friston et al., 2003). The empirical work in this paper shows that a minimal model using only four parameters is able to provide a good fit to our particular ECOG data set.

In the HB model, synchronization of spike timing is achieved through balanced excitation and inhibition of ensembles of integrate and fire cells. This is a perfectly plausible mechanism and may indeed be an accurate description of how cells in the brain synchronize. As we have described in the introduction there are multiple alternative ways that cells might synchronize. The WCO-based recognition-by-synchronization module might be seen as

an improvement on the HB module in two ways. First, it is less committed to a specific biophysical synchronization mechanism, which we see to be an advantage because the mechanism by which neurons in the mammalian auditory synchronize is currently unknown. Second, when used as a forward model of neuroimaging data it is computationally more efficient because a larger integration step size can be used.

The TS mechanism we have investigated is similar to other models of neural processing that rely on transient dynamics (Rabinovich, Huerta, & Laurent, 2008). The Liquid State Machine (LSM) (Maass, Natschlager, & Markram, 2002), for example, uses OT features and the temporal embedding idea proposed in the HB model, but then applies standard methods for recognizing the resulting static patterns. This results in good pattern discrimination abilities (Verstraeten et al., 2005), though not as accurate as a recent approach based on OT features (Gutig & Sompolinsky, 2009). Further, LSMs do not generate a gamma burst as an integral part of the recognition process, so would not be so appropriate as a forward model for the sort of neuroimaging data addressed here.

The notion that regions higher up in the auditory cortical hierarchy process information at longer time scales has recently been made use of in a model of auditory sequence recognition based on stable heteroclinic channels (Kiebel, Kriegstein, Daunizeau, & Friston, 2009). Moreover, the approach developed in that work derives from a Bayesian perspective in which cortical hierarchies embody a generative model which is then inverted during the pattern recognition process. Generative models of speech production are, as yet however, still in the early stages of development. This currently limits the ecological validity of such a generative modelling approach.

The importance of a hierarchy of temporal scales is emphasized in recent work by Ghitza (2011) who provides evidence that current models of speech perception, which are driven by acoustic features alone, provide an incomplete description of speech recognition phenomena. An alternative description which highlights the role of decoding time provides a better match to human perceptual performance, and they suggest that decoding time is governed by a cascade of neural oscillators operating at different time scales.

Finally, we have used the dynamic pattern recognition model to predict activity in only a single brain region, the posterior superior temporal sulcus. This region has been identified in several fMRI paradigms where normal speech is compared with various unintelligible speech foils (Leff et al., 2008; Obleser, Wise, Dresner, & Scott, 2007; Vouloumanos, Kiehl, Werker, & Liddle, 2001). Given that task-dependent BOLD responses and gamma oscillations are coupled, it seems reasonable to suggest that the activations reported in these studies could be driven by the gamma burst predicted to occur when spike time synchronization occurs. This response will be more robust for sensory items with long-term neural representations maintained by repeated exposure (i.e., words), compared with acoustically comparable items that do not (i.e., non-words).

Whilst the work in this paper provides a useful starting point, it does not make use of the network view of brain function; Price, Thierry, and Griffiths (2005), for example, propose that the human brain has not developed macro-anatomical structures dedicated to speech processing, but rather that speech-specific processing emerges at the level of functional connectivity among distributed regions. The ECoG data we have analysed has recordings of activity from 64 electrodes placed over fronto-temporal cortex, yet we have modelled data from only a single electrode. Extension of our modelling to include multiple regions, as in Corchs and Deco (2004); Husain et al. (2004), is therefore an important direction for future work.

Acknowledgement

This work was supported by the Wellcome Trust.

Appendix A. Mel frequency cepstral coefficients

The MFCCs were derived using a standard data processing pipeline (Gutig & Sompolinsky, 2009). As the magnitude spectrum of speech is known to be stationary over a period of approximately $t_{\text{win}} = 20$–100 ms (Rabiner & Juang, 1993) we broke up each speech time series into frames of length $t_{\text{win}} = 50$ ms. These frames were offset from each other by $t_{\text{off}} = 50$ ms, resulting in non-overlapping frames. As the overall time series was $t_{\text{fix}} = 900$ ms, this resulted in 18 frames. The frame size and offset are a little larger than is standard but these settings were found to give perfect speech recognition performance on non-noisy data (see main text).

For each frame of data we first applied a pre-emphasis filter

$$\hat{x}_n = x_n - \alpha x_{n-1} \tag{16}$$

with $\alpha = 0.97$. We then applied a 512 point FFT to the time series $\hat{x}_n$ to obtain the power spectrum p_x. This was then multiplied by a set of $k = 1..K$ filters (see below), s_k, to give

$$z_k = s_k p_x. \tag{17}$$

The MFCCs were then computed as

$$a = \text{DCT}(\log(z)) \tag{18}$$

where DCT is the Discrete Cosine Transform. Thus, the MFCCs are given by a cosine transform of the logarithm of the filtered power spectra. The filters are triangular-shaped and spaced uniformly on the 'mel-frequency scale'

$$f_m = 2595 \log_{10}(1 + f/700) \tag{19}$$

where f is the frequency in Hz and f_m is the frequency in 'mels'. The mel scale has been designed such that intervals are equally separated in perceptual distance, with $f_m = 1000$ mels being equivalent to $f = 1000$ Hz. The relationship is linear below 1000 Hz and sub-linear above it.

The cepstral parameters were then weighted by multiplying the kth parameter, a_k, by

$$w(k) = 1 + \frac{K}{2} \sin\left(\frac{k}{K}\pi\right) \tag{20}$$

for $k = 1..K$, a procedure known as 'liftering'. The first and second order derivatives of the MFCCs (with respect to time) are sometimes used as additional features (Gutig & Sompolinsky, 2009) but did not improve recognition performance on our database.

Appendix B. Frequency dependent coupling

The main text describes a WCO–TS model based on uniform coupling, $w_{kk'} = A$. We also allowed for an augmented model in which coupling strength also depends on the mean frequency of the ensemble,

$$\bar{f}(x, t) = \frac{1}{K} \sum_{k=1}^{K} f_k(x, t). \tag{21}$$

This takes the sigmoidal form

$$w_{kk'} = \frac{A}{1 + \exp(-a)} \tag{22}$$

$$a = \beta_1(\bar{f}(x, t) - \beta_2 f_b).$$

The purpose of this frequency-dependence is to reduce synchronization at lower frequencies. The possibility that synchronization increases with frequency is commensurate with in-vitro cell recordings (Rocha, Doiron, Shea-Brown, Josic, & Reyes, 2007) and computer simulation of both integrate and fire and Hodgkin–Huxley type models (Chawla, Lumer, & Friston, 1999).

References

Abbott, L. (2001). The timing game. *Nature Neuroscience*, *4*(2), 115–116.

Bartos, M., Vida, I., & Jonas, P. (2007). Synaptic mechanisms of synchronized gamma oscillations in inhibitory interneuron networks. *Nature Reviews Neuroscience*, *8*(1), 45–56.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., et al. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex*, *10*(5), 512–528.

Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford: Oxford University Press.

Borisyuk, R. M., & Kazanovich, Y. B. (2004). Oscillatory model of attention-guided object selection and novelty detection. *Neural Networks*, *17*(7), 899–915.

Brown, E., Moehlis, J., & Holmes, P. (2004). On the phase reduction and response dynamics of neural oscillator populations. *Neural Computation*, *16*(4), 673–715.

Brunel, N., & Hakim, V. (1999). Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural Computation*, *11*(7), 1621–1671.

Brunel, N., & Wang, X. J. (2003). What determines the frequency of fast network oscillations with irregular neural discharges? I. Synaptic dynamics and excitation–inhibition balance. *Journal of Neurophysiology*, *90*(1), 415–430.

Buzsaki, G. (2006). *Rhythms of the brain*. Oxford University Press.

Canolty, R. T., Soltani, M., Dalal, S. S., Edwards, E., Dronkers, N. F., Nagarajan, S. S., et al. (2007). Spatiotemporal dynamics of word processing in the human brain. *Frontiers in Neuroscience*, *1*(1), 185–196.

Casseday, J., Fremouw, T., & Covey, E. (2002). In D. Oertel, R. Fay, & A. Popper (Eds.), *Integrative functions in the mammalian auditory pathway* (pp. 238–318). Springer.

Chawla, D., Lumer, E. D., & Friston, K. J. (1999). The relationship between synchronization among neuronal populations and their mean activity levels. *Neural Computation*, *11*(6), 1389–1411.

Corchs, S., & Deco, G. (2001). A neurodynamical model for selective visual attention using oscillators. *Neural Networks*, *14*(8), 981–990.

Corchs, S., & Deco, G. (2004). Feature-based attention in human visual cortex: simulation of fMRI data. *NeuroImage*, *21*(1), 36–45.

David, O., & Friston, K. (2003). A neural mass model for MEG/EEG: coupling and neuronal dynamics. *NeuroImage*, *20*(3), 1743–1755.

Duda, R., & Hart, P. (1973). *Pattern classification and scene analysis*. John Wiley and Sons.

Elliott, T., & Theunissen, F. (2009). The modulation transfer function for speech intelligibility. *PLoS Computational Biology*, *5*(3), e1000302.

Ermentrout, G. B., & Kleinfeld, D. (2001). Traveling electrical waves in cortex: insights from phase dynamics and speculation on a computational role. *Neuron*, *29*(1), 33–44.

Eulitz, C., Maess, B., Pantev, C., Friederici, A. D., Feige, B., & Elbert, T. (1996). Oscillatory neuromagnetic activity induced by language and non-language stimuli. *Brain Research Cognitive Brain Research*, *4*(2), 121–132.

Fries, P. (2005). A mechanism for cognitive dynamics. *Trends in Cognitive Sciences*, *9*(10), 474–480.

Friston, K., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, *19*(4), 1273–1302.

Geisler, C., Brunel, N., & Wang, X. J. (2005). Contributions of intrinsic membrane dynamics to fast network oscillations with irregular neuronal discharges. *Journal of Neurophysiology*, *94*(6), 4344–4361.

Gelman, A., Carlin, J., Stern, H., & Rubin, D. (1995). *Bayesian data analysis*. Boca Raton: Chapman and Hall.

Ghitza, O. (1986). Auditory nerve representation as a front-end for speech recognition in a noisy environment. *Computer Speech and Language*, *1*, 109–130.

Ghitza, O. (2007). Using auditory feedback and rhythmicity for diphone discrimination of degraded speech. In *Proceed. intern. conf. on phonetics*. August (pp. 163–168).

Ghitza, O. (2011). Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. *Frontiers in Psychology*, *2*, 130. Disponible sur http://dx.doi.org/10.3389/fpsyg.2011.00130.

Girolami, M. (2008). Bayesian inference for differential equations. *Theoretical Computer Science*, *408*(1), 4–16.

Goense, J. M., & Logothetis, N. K. (2008). Neurophysiology of the BOLD fMRI signal in awake monkeys. *Current Biology*, *18*(9), 631–640.

Gutig, R., & Sompolinsky, H. (2009). Time-warp-invariant neuronal processing. *PLoS Biology*, *7*(7), e1000141.

Gutkin, B. S., Ermentrout, G. B., & Reyes, A. D. (2005). Phase-response curves give the responses of neurons to transient inputs. *Journal of Neurophysiology*, *94*(2), 1623–1635.

Hansel, D., Mato, G., & Meunier, C. (1995). Synchrony in excitatory neural networks. *Neural Computation*, *7*(2), 307–337.

Harrison, R., Harel, N., Panesar, J., Mori, N., & Mount, R. (2000). Local haemodynamic changes associated with neural activity in auditory cortex. *Acta Otolaryngol*, *120*, 255–258.

Hopfield, J. J. (2004). Encoding for computation: recognizing brief dynamical patterns by exploiting effects of weak rhythms on action-potential timing. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(16), 6255–6260. Disponible sur http://dx.doi.org/10.1073/pnas.0401125101.

Hopfield, J., & Brody, C. (2000). What is a moment? cortical sensory integration over a brief interval. *Proceedings of the National Academy of Sciences*, *97*(25), 13919–13924.

Hopfield, J., & Brody, C. (2001). What is a moment? transient synchrony as a collective mechanism for spatiotemporal integration. *Proceedings of the National Academy of Sciences*, *98*(3), 1282–1287.

Hoppensteadt, F., & Izhikevich, E. (1997). *Weakly connected neural networks*. New York, USA: Springer Verlag.

Hromadka, T., Deweese, M., & Zador, A. (2008). Sparse representation of sounds in the unanesthetized auditory cortex. *PLoS Biology*, *6*(1), e16.

Husain, F. T., Tagamets, M. A., Fromm, S. J., Braun, A. R., & Horwitz, B. (2004). Relating neuronal dynamics for auditory object processing to neuroimaging activity: a computational modeling and an fMRI study. *NeuroImage*, *21*(4), 1701–1720.

Jensen, O., Kaiser, J., & Lachaux, J. (2007). Human gamma-frequency oscillations associated with attention and memory. *Trends in Neurosciences*, *30*(7), 317–324.

Kiebel, S., Kriegstein, K., Daunizeau, J., & Friston, K. (2009). Recognizing sequences of sequences. *PLoS Computational Biology*, *5*(8), e1000464.

Kloeden, P., & Platen, E. (1999). *Numerical solution of stochastic differential equations*. Berlin: Springer.

Kuramoto, Y. (1984). *Chemical oscillations, waves and turbulence*. New York, USA: Springer Verlag.

Lee, C., Glass, J., & Ghitza, O. (2011). An efferent-inspired auditory model front-end for speech recognition. In *Proceedings of interspeech, florence*.

Lee, S., Sen, K., & Kopell, N. (2009). Cortical gamma rhythms modulate NMDAR-mediated spike timing dependent plasticity in a biophysical model. *PLoS Computational Biology*, *5*(12), e1000602.

Leff, A., Schofield, T., Stephan, K., Crinion, J., Friston, K., & Price, C. (2008). The cortical dynamics of intelligible speech. *Journal of Neuroscience*, *28*(49), 13209–13215.

Loiselle, S., Rouat, J., Pressnitzer, D., & Thorpe, S. (2005). Exploration of rank order coding with spiking neural networks for speech recognition. In *Proceedings of the international joint conference on neural networks. Montreal, Canada*.

Lutzenberger, W., Pulvermuller, F., & Birbaumer, N. (1994). Words and pseudowords elicit distinct patterns of 30-Hz EEG responses in humans. *Neuroscience Letters*, *176*(1), 115–118.

Maass, W., Natschlager, T., & Markram, H. (2002). Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Computation*, *14*(11), 2531–2560.

Marr, D., & Poggio, T. (1976). From understanding computation to understanding neural circuitry (Rapport technique). MIT artifical intelligence laboratory. (Available from: http://computer-refuge.org/bitsavers/pdf/mit/ai/aim/AIM-357.pdf).

May, P., & Tiitinen, H. (2007). The role of adaptation-based memory in auditory cortex. *International Congress Series*, *1300*, 53–56. (New frontiers in biomagnetism. Proceedings of the 15th international conference on biomagnetism. Vancouver, BC, Canada. August 21–25, 2006).

Miller, J. L., Grosjean, F., & Lomanto, C. (1984). Articulation rate and its variability in spontaneous speech: a reanalysis and some implications. *Phonetica*, *41*(4), 215–225.

Mitra, P., & Pesaran, B. (1999). Analysis of dynamic brain imaging data. *Biophysical Journal*, *76*, 691–708.

Niebur, E., & Koch, C. (1994). A model for the neuronal implementation of selective visual attention based on temporal correlation among neurons. *Journal of Computational Neuroscience*, *1*, 141–158.

Obleser, J., Wise, R., Dresner, M. A., & Scott, S. (2007). Functional integration across brain regions improves speech perception under adverse listening conditions. *Journal of Neuroscience*, *27*, 2283–2289.

Pearce, D., & Hirsch, H. (2000). The Aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions. In *6th international conference on spoken language processing. Beijing, China*.

Penny, W., Litvak, V., Fuentemilla, L., Duzel, E., & Friston, K. (2009). Dynamic causal models for phase coupling. *Journal of Neuroscience Methods*, *183*(1), 19–30.

Penny, W., Stephan, K., Mechelli, A., & Friston, K. (2004). Comparing dynamic causal models. *NeuroImage*, *22*(3), 1157–1172.

Pfeuty, B., Mato, G., Golomb, D., & Hansel, D. (2003). Electrical synapses and synchrony: the role of intrinsic currents. *Journal of Neuroscience*, *23*(15), 6280–6294.

Price, C., Thierry, G., & Griffiths, T. (2005). Speech-specific auditory processing: where is it? *Trends in Cognitive Sciences*, *9*(6), 271–276.

Pulvermuller, F., Eulitz, C., Pantev, C., Mohr, B., Feige, B., Lutzenberger, W., et al. (1996). High-frequency cortical responses reflect lexical processing: an MEG study. *Electroencephalography and Clinical Neurophysiology*, *98*(1), 76–85.

Rabiner, L., & Juang, B. (1993). *Fundamentals of speech recognition*. Prentice-Hall.

Rabinovich, M., Huerta, R., & Laurent, G. (2008). Transient dynamics for neural processing. *Science*, *321*(5885), 48–50.

Rocha, J. de la, Doiron, B., Shea-Brown, E., Josic, K., & Reyes, A. (2007). Correlation between neural spike trains increases with firing rate. *Nature*, *448*(7155), 802–806.

Rouat, J., Loiselle, S., & Molotchnikoff, S. (2011). Variable frame rate hierarchical analysis for robust speech recognition. In *Proceedings IEEE international conference on intelligent robots and systems*.

Roweis, S. (1998). Speech processing background (Rapport technique). New York University. (Available from: www.cs.nyu.edu/~roweis/notes/spba4.ps.gz).

Roweis, S. (2009). Adding up oscillations (Rapport technique). New York University. (Available from: www.cs.nyu.edu/~roweis/notes/sinwslides.ps.gz).

Shamir, M., Ghitza, O., Epstein, S., & Kopell, N. (2009). Representation of time-varying stimuli by a network exhibiting oscillations on a faster time scale. *PLoS Computational Biology*, *5*(5), e1000370. Disponible sur http://dx.doi.org/10.1371/journal.pcbi.1000370.

Singer, W. (1999). Neuronal synchrony: a versatile code for the definition of relations? *Neuron*, *24*(1), 49–65. 111–125.

Tiesinga, P. H., & Jose, J. V. (2000). Robust gamma oscillations in networks of inhibitory hippocampal interneurons. *Network*, *11*(1), 1–23.

Traub, R. D., Schmitz, D., Jefferys, J. G., & Draguhn, A. (1999). High-frequency population oscillations are predicted to occur in hippocampal pyramidal neuronal networks interconnected by axoaxonal gap junctions. *Neuroscience*, *92*(2), 407–426.

Tsodyks, M., Mitkov, I., & Sompolinsky, H. (1993). Pattern of synchrony in inhomogeneous networks of oscillators with pulse interactions. *Physical Review Letters*, *71*, 1280–1283.

Ulanovsky, N., Las, L., Farkas, D., & Nelken, I. (2004). Multiple time scales of adaptation in auditory cortex neurons. *Journal of Neuroscience*, *24*, 10440–10453.

Ursino, M., Cona, F., & Zavaglia, M. (2010). The generation of rhythms within a cortical region: analysis of a neural mass model. *NeuroImage*, *52*(3), 1080–1094.

Varela, F., Lachaux, J., Rodriguez, E., & Martinerie, J. (2001). The brainweb: phase synchronization and large-scale integration. *Nature Reviews Neuroscience*, *2*(4), 229–239.

Verstraeten, D., Schrauwen, B., Stroobandt, D., & Campenhout, J. V. (2005). Isolated word recognition with the liquid state machine: a case study. *Information Processing Letters*, *95*, 521–528.

Vouloumanos, A., Kiehl, K., Werker, J., & Liddle, P. (2001). Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech. *Journal of Cognitive Neuroscience*, *13*(7), 994–1005.

Vreeswijk, C. V., Abbott, L. F., & Ermentrout, G. B. (1994). When inhibition not excitation synchronizes neural firing. *Journal of Computational Neuroscience*, *1*(4), 313–321.

Walker, W., Lamere, P., Kwok, P., Raj, B., & Singh, R. (2004). Sphinx-4: a flexible open source framework for speech recognition. Technical Report No. SMLI TR-2004-139, Sun Microsystems Laboratories.

Wang, X. J. (2010). Neurophysiological and computational principles of cortical rhythms in cognition. *Physiological Reviews*, *90*(3), 1195–1268. Disponible sur http://dx.doi.org/10.1152/physrev.00035.2008.

Wang, X. J., & Buzsaki, G. (1996). Gamma oscillation by synaptic inhibition in a hippocampal interneuronal network model. *Journal of Neuroscience*, *16*(20), 6402–6413.

Wang, D., & Terman, D. (1997). Image segmentation based on oscillatory correlation. *Neural Computation*, *9*, 805–836.

White, J. A., Chow, C. C., Ritt, J., Soto-Trevino, C., & Kopell, N. (1998). Synchronization and oscillatory dynamics in heterogeneous, mutually inhibited neurons. *Journal of Computational Neuroscience*, *5*(1), 5–16.

Whittington, M. A., Traub, R. D., Kopell, N., Ermentrout, B., & Buhl, E. H. (2000). Inhibition-based rhythms: experimental and mathematical observations on network dynamics. *International Journal of Psychophysiology*, *38*(3), 315–336.

Woodland, P. (2003). Htk3. Manuel de logiciel. (Available from: http://htk.eng.cam.ac.uk/).

Zhang, F., Miller, C., Robinson, B., Abbas, P., & Hu, N. (2007). Changes across time in spike rate and spike amplitude of auditory nerve fibers stimulated by electric pulse trains. *Journal of the Association for Research in Otolaryngology*, *8*(3), 356–372.