ARE NEURAL OSCILLATIONS THE SUBSTRATE OF AUDITORY GROUPING?

Guy J. Brown[1], Martin Cooke[1] and Eric Mousset[1,2]

[1]*Department of Computer Science, University of Sheffield,*
Regent Court, 211 Portobello Street, Sheffield S1 4DP, United Kingdom

[2]*INGENIA, 92 bis, Avenue Victor Cresson, 92130 Issy-les-Moulineaux, France*

Email: {g.brown,m.cooke}@dcs.shef.ac.uk, mousset@ingenia.fr

Published in: Proceedings of the ESCA Tutorial and Workshop on the Auditory Basis of Speech Perception, Keele University, July 15-19, 1996

ABSTRACT

How are acoustic features that are extracted in remote regions of the auditory system bound together to form a perceptual whole? We consider the evidence for a solution to this so-called *binding problem*, which proposes that the responses of feature detecting cells are bound together by the synchronisation of oscillatory firing activity. Four models of auditory grouping based on neural oscillators are reviewed, and issues arising from these models are discussed.

1. INTRODUCTION

Recent advances in auditory neuroscience support the notion that different properties of acoustic events (such as periodicity, spatial location and spectral shape) are extracted at separate locations in the auditory system [22]. Nonetheless, we perceive auditory events as meaningful wholes, not as parts. In other words, the auditory system is able to bind together features represented in remote neural structures to form perceptual wholes. The mechanism of this binding process is the subject of our paper.

Clearly, the binding problem does not arise only in auditory perception. Other sensory channels, such as the visual system, must also combine fragmentary representations of stimuli from separate neural structures. Further, features must be integrated across different sensory and motor systems, both in perception and recall. For example, audiovisual integration is apparent in the well-known McGurk effect [21], in which the perception of speech sounds is influenced by the image of the speaker's face. Such integration of information from different sensory modalities suggests that the brain employs a universal neural mechanism for feature binding.

The traditional solution to the binding problem invokes a hierarchy of increasingly specialised feature detecting cells. It has been hypothesised that 'cardinal' cells at the highest level of this hierarchy might be tuned to detect the appearance of particular visual objects [3]. However, neuroanatomical and neuropsychological studies have raised so many objections to this approach that it must now be regarded as untenable (see [9] for a review). In particular, the theory requires the existence of cortical sites which receive projections from all the neural regions specialised for processing the different sensory attributes of a stimulus. Despite intensive study of cortical structure, such sites have not been identified. Also, it is likely that the number of 'cardinal' neurons required to represent the almost limitless variety of sensory stimuli is prohibitively large.

A different solution to the binding problem has been proposed by von der Malsburg [19,20]. He suggests that the responses of feature detecting cells may be bound together by the temporal synchronisation of their firing activity. Those neurons representing features of the same perceptual event would be synchronised, whereas neurons representing features of different events would be desynchronised. Evidence supporting this so-called *correlation theory* has come from studies of the visual and olfactory systems, which report that stimuli evoke synchronised neural oscillations in functionally related (but remote) areas of the cortex [10]. Additionally, synchronised oscillations evoked by acoustic stimuli have been observed in the auditory cortex [16]. These experimental findings have led to more specific formulations of von der Malsburg's scheme, which propose that feature binding is signalled by the phase of neural oscillations (e.g., [10,20,26]).

Until recently, the role of neural oscillations in feature binding had been studied predominantly in the visual and olfactory systems. Hence, many of the issues in this area might be unfamiliar to members of the hearing research community. We aim to address this deficiency in the remainder of this article, which contains elements of tutorial, review and position paper. First, we review the evidence that neural oscillations have a functional role in the auditory system, and consider the origin of such oscillations. We then review four recent auditory models which employ oscillatory neural networks, and draw on this literature to raise some key issues in the field. Finally, we discuss some outstanding challenges that theories of feature binding must address.

1.1. Evidence for oscillations in the auditory system

Evidence for synchronised neural oscillations in the neocortex has come from studies of the electroencephalogram (EEG), and from field potentials recorded with intracerebral electrodes (see [23] for a review). These oscillations usually lie in the β and γ range of the EEG (15-60 Hz), and are referred to as '40 Hz' oscillations. In the auditory domain, 40 Hz oscillations have been observed in EEGs evoked by tonal stimuli [13]. Also, a recent study in which alert human subjects were presented with pairs of clicks has provided direct evidence for the role of neural oscillations in auditory grouping [16]. For click pairs presented less than 12-15 ms apart, subjects reported a single source and simultaneous magneto-encephalography (MEG) recordings from the auditory cortex showed a single 40 Hz oscillation. The perceived timbre in this condition was different to that of an isolated click, indicating that the two clicks had been fused into a single percept. For interstimulus intervals greater than 12-15 ms, listeners reported two clicks and MEG recordings showed a second 40 Hz oscillation. These findings support the view that 40 Hz oscillatory activity in the auditory cortex is correlated with temporal grouping of auditory stimuli.

It should be stressed that 40 Hz oscillations implicated in auditory feature binding are not phase-locked to the stimulating waveform. They should therefore be distinguished from other oscillatory responses in the auditory system, such as those of 'chopper' cells in the dorsal cochlear nucleus. Chopper cells oscillate at a much higher frequency (up to 500 Hz) and tend to phase-lock to the envelope of amplitude modulated stimuli [12].

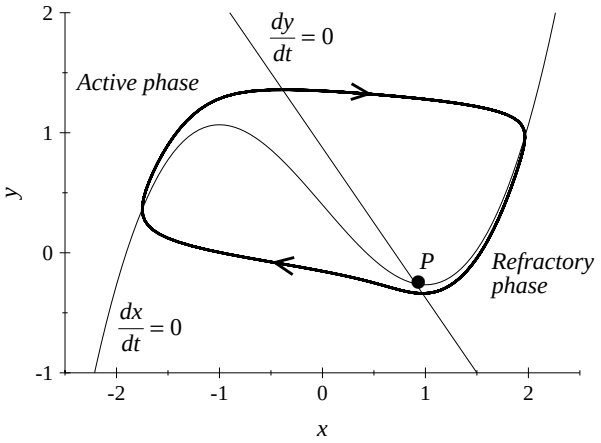


Figure 1: Behaviour of the FitzHugh equations in the phase plane, with parameters a=0.7, b=0.8, c=3 and I=-0.4. The periodic solution (limit cycle) is shown in bold.

1.2. The origin of neural oscillations: A dynamical systems perspective

Oscillatory firing activity usually originates from the interaction of mutually connected excitatory and inhibitory processes. Such interactions may occur within single neurons due to the coupling of excitatory and inhibitory membrane conductances, or may occur in networks of excitatory and inhibitory neurons with feedback connections.

Considerable insight into the origin of neural oscillations can be gained by viewing neurons (and neural networks) as dynamical systems. For example, consider the caricature of the Hodgkin-Huxley equations proposed by FitzHugh [11] (the so-called Bonhoeffer-van der Pol model), which is described by the following coupled nonlinear differential equations:

$$\frac{dx}{dt} = c\left(x + y - x^3/3 + I\right) \tag{1a}$$

$$\frac{dy}{dt} = -\frac{1}{c}(x - a + by) \tag{1b}$$

Here, x and y are the dynamical variables, I is the external input and a, b and c are constants. The system (1) is usually interpreted as a model of action potential generation in a single neuron, in which x corresponds to membrane potential and y is related to refractoriness. Alternatively, (1) may be interpreted as a mean-field approximation to a network of coupled excitatory and inhibitory neurons. In this case, x and y represent the average firing activity of a group of excitatory neurons and a group of inhibitory neurons respectively.

The instantaneous state of the system (1) is given by the pair of values (x,y). Solutions of the system can therefore be viewed as trajectories in the (x,y) plane, which is called the *phase plane*. Among the most important characteristics of any dynamical system are its states of equilibrium, which correspond to so-called *fixed points* in the phase plane. These can be illustrated by plotting the *nullclines* of the system. The x-nullcline is the locus of points in the phase plane along which dx/dt=0. Similarly, the y-nullcline is the locus of points along which dy/dt=0. Fixed points occur at the intersection of the x- and y-nullclines (i.e., when dx/dt=0 and dy/dt=0).

The phase plane for the FitzHugh equations (1) is shown in Figure 1. The x-nullcline (a cubic curve) and the y-nullcline (a line of slope -1/b) intersect at a single fixed point P. This point corresponds to the resting state of the neuron (or network). In response to an instantaneous input of sufficient magnitude, the state of the system moves rapidly from P to a region of negative x, and then passes through a region of positive x back to P.

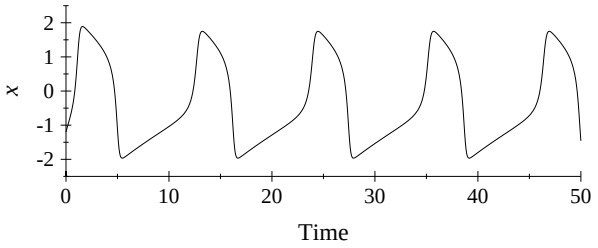


Figure 2: Behaviour of the variable x in the FitzHugh equations for the limit cycle solution shown in Figure 1. If x is interpreted as membrane potential, the output of the system resembles a periodic train of action potentials.

Interpreting x as membrane potential, this corresponds to a phase of firing activity followed by a phase of refractoriness: in other words, a single action potential. Different behaviour is observed when the input is sustained. In this case, (1) has a periodic solution (a so-called *limit cycle* oscillation), shown by the bold line in Figure 1. Plotting x as membrane potential (Figure 2) reveals a periodic train of action potentials: a neural oscillation. Note that the state of the system changes relatively slowly in the active and refractory phases, but moves from one phase to the other relatively quickly. Such alternating phases of fast and slow change are characteristic of so-called *relaxation oscillators.* Oscillators of this type are the building blocks of two of the auditory models described in the following section.

2. NEURAL OSCILLATOR MODELS OF AUDITORY GROUPING

There is a close relationship between the binding problem and so-called 'auditory grouping' [5]; the latter concerns the issue of *why* particular features are combined to form perceptual wholes, whereas the binding problem concerns the issue of *how* such groups of features are represented in the brain. However, very few computational models of auditory grouping have also addressed the neural mechanisms of feature binding. The following section reviews four modelling studies which do address both processes, and have done so using architectures based on neural oscillators.

2.1. von der Malsburg & Schneider

The model of auditory grouping described by von der Malsburg & Schneider [20] is a form of Malsburg's correlation theory of brain function (see Section 1). It employs a neural architecture in which each member of a fully-connected network of excitatory cells (E-cells) receives an input from one frequency band of the auditory periphery. In addition, all E-cells receive inhibition from a common inhibitory cell (H-cell). The model is described by the following difference equations:

$$E_i(t+\Delta t) = N_i(t) \cdot L\left[\sum_{i<>j} s_{ij}(t)E_j(t) + I_i(t) + \alpha E_i(t) - s_{he}H(t) + \eta_i(t)\right] \tag{2}$$

$$H(t+\Delta t) = L\left[\beta H(t) + s_{eh}\sum_j E_j(t)\right] \tag{3}$$

Here, $E_i(t)$ represents the activity of the ith E-cell at time t, $H(t)$ represents the activity of the inhibitory cell, $\eta_i(t)$ is a noise term and α and β are parameters. The external input to the ith E-cell is $I_i(t)$. The term s_{ij} determines the coupling strength between pairs of E-cells, whereas s_{he} and s_{eh} determine the strength of reciprocal connections between each E-cell and the H-cell. The function $N_i(t)$ is 1 during bursts of firing activity and 0 during refractory periods. The nonlinearity $L(x)$ constrains the output

to a value between zero and unity.

The model assumes that coupling strengths between pairs of E-cells are modified on a fast timescale, according to their state of synchronisation. E-cells which receive simultaneous inputs tend to become synchronised by the excitatory links between them, and tend to become desynchronised from other cells due to the influence of inhibition from the H-cell. The network therefore displays a sensitivity to the common onset of acoustic components. However, the model is limited to this single grouping principle; no information about the distance between acoustic components in time and frequency is preserved, and hence the model is unable to reproduce other well-known aspects of auditory organisation such as grouping by temporal and frequency proximity.

2.2. Wang

Wang [26] has recently described a model of auditory grouping in which time-frequency patterns are represented on a two-dimensional grid of relaxation oscillators. It is hypothesised that the time axis of this grid is created by a system of delay lines. Individual oscillators have the form

$$\frac{dx}{dt} = 3x - x^3 + 2 - y + I + S + \eta \tag{4a}$$

$$\frac{dy}{dt} = \varepsilon[\gamma(1 + \tanh(x/\beta)) - y] \tag{4b}$$

where ε, γ and β are parameters, S represents the coupling from other oscillators in the network, and η is a noise term. Oscillators are connected to one another by two excitatory links, one of constant weight and one of variable weight. The constant weight between a pair of oscillators on the grid falls off exponentially with the distance between them. This endows the oscillator network with sensitivity to the frequency and temporal proximity of acoustic components. The variable ('dynamic') weights are modified during simulation, depending on the state of synchronisation in the network. Additionally, all oscillators receive inhibition from a global inhibitor.

When presented with an input, Wang's oscillator network rapidly achieves synchronisation of related groups of features through local excitatory connections, and desynchronisation of different groups of features through global inhibition. The model is able to replicate several auditory grouping phenomena that occur in the perception of repeating tone sequences, such as sequential capturing and the competition between alternative organisations (see [5]).

2.3. Brown & Cooke

A four-stage neural oscillator model of auditory grouping has been proposed by Brown & Cooke [6]. In the first stage of their model, peripheral auditory processing is simulated by a bank of bandpass filters and a model of inner hair cell function. In the second stage, simulated auditory nerve firing patterns, derived from the hair cell model, are processed by an array of onset cells. Thirdly, activity in the onset map is used to modify the coupling strengths between neurons in a fully connected neural network. Specifically, the neural network model proposed by Bauer & Martienssen [4] is adopted, in which the phase dynamics of each neural oscillator is represented by a sine circle map. The sine circle map $\varphi(x)$ is given by

$$\varphi(x) = x + \Omega + \frac{k}{2\pi}\sin(2\pi x) + \eta \pmod 1 \tag{5}$$

where η is a noise term. The new phase $\theta(t+1)$ of a neuron is computed by applying the circle map (5) on the old phase $\theta(t)$ and on an input value υ, weighted with a coupling strength κ:

$$\theta(t+1) = \frac{1}{1+\kappa}[\varphi(\theta(t)) + \kappa\varphi(\upsilon(t))] \tag{6}$$

Weights are updated during simulation by reducing the coupling strength between filter channels whose onset cells do not exhibit the same level of activity at the same time. The final stage of the model is an attentional mechanism, motivated by Crick's [8] hypothesis that an attentional 'searchlight' is located in the thalamus. Despite its simplicity, the model closely matches the performance of human listeners in two-tone streaming studies. In particular, it is able to account for grouping by temporal and frequency proximity, common onset and good continuation.

2.4. Liu, Yamaguchi and Shimizu

Strictly, the model described by Liu *et al*. [18] addresses vowel recognition rather than auditory grouping. However, it is considered here because it employs neural oscillators, and also because its recognition architecture may be interpreted as a mechanism of top-down (schema-driven) auditory grouping [5].

The model consists of an input layer and three neural layers, which are referred to as the A, B and C centres. The input is a time series of mel-scaled linear prediction coefficients (LPC). Each oscillator in the A and B centres has the form

$$\frac{dx}{dt} = -x + \alpha s(x) + \beta s(y) + c_1 \tag{7a}$$

$$\frac{dy}{dt} = -y + \gamma s(x) + c_2 \tag{7b}$$

where

$$s(x) = 1/2[\tanh(x/\lambda) + 1] \tag{8}$$

and α, β, γ, λ, c_1 and c_2 are parameters. At each epoch, the A centre identifies the local peaks in the input pattern and encodes them as groups of independent oscillations. It is assumed that these groups correspond to vowel formants. The B centre constitutes a simple associative memory, in which connections are hard-wired to reflect the variation of Japanese vowel formant frequencies. Associative interactions within the B centre, together with reciprocal top-down and bottom-up interactions between the A and B centres, lead to the activation of a vowel as a global pattern of synchronised oscillations. The C centre evaluates the synchronisation in each of three formant regions, and outputs a vowel category.

Liu *et al*. demonstrate that the top-down connections in their model are essential for the synchronisation of the A and B centres. Also, the top-down connections from the B centre confer robustness in noise; the model is able to recognise vowels in a background of multi-speaker babble.

3. ISSUES ARISING FROM THE MODELS

A consideration of the similarities and differences between the models reviewed above provides some insights into the neural oscillator theory. We discuss a number of these issues below.

3.1. Classes of oscillator models

The relaxation oscillators employed in the Wang and Liu *et al*. models bear some resemblance to the FitzHugh equations presented in equation (1). However, by using sigmoidal functions for the *y*-nullcline, both models provide extra flexibility which is not present in the FitzHugh equations. For example, the parameter γ in the Wang model (equation 4b) allows the relative time that an oscillator spends in the active phase and refractory phase to be determined [26].

The dynamics of the Malsburg & Schneider model cannot be directly compared with relaxation oscillator models because it is presented as a system of difference equations, rather than differential equations. However, the structure of the models are similar; like a relaxation oscillator, the Malsburg & Schneider model consists of mutually coupled excitatory and inhibitory processes. Indeed, the model of Wang is very close in concept

to that of Malsburg & Schneider; both consist of coupled excitatory cells and a global inhibitor (see also Section 3.2). However, the formulation of Malsburg & Schneider's oscillator model is perhaps rather *ad hoc*; Wang's has a similar structure, but with dynamics that are closely based on those of biological neurons.

The model of Brown & Cooke, based on Bauer & Martienssen's chaotic oscillators [4], differs significantly from the other approaches; rather than creating oscillations through the interaction of excitatory and inhibitory mechanisms, it models each oscillator with a single phase variable. The chaotic oscillations produced by this model allow a large number of groups to be represented, but have a considerable disadvantage; a cross-correlation analysis is needed in order to evaluate the synchronisation of the network [6]. In contrast, relaxation oscillators exhibit rapid transitions between active and refractory phases (see Figure 2), and groups of simultaneously active neurons can therefore be identified by applying a simple threshold.

A common feature in the oscillator models of Section 2 is the inclusion of a noise term. The purpose of the noise is to assist the desynchronisation of groups of oscillators which happen to start from very similar initial conditions (so-called 'symmetry breaking' [20,26]). Also, the addition of noise allows the robustness of a model to be evaluated [26].

3.2. Local connectivity vs. global connectivity

The models described in Section 2 differ in the connectivity of their oscillator networks; two models employ global (all-to-all) connectivity (Malsburg & Schneider, Brown & Cooke), whereas two employ lateral coupling over a limited range of frequencies (Wang, Liu *et al*.). The use of lateral connections suggests an influence of earlier neural oscillator models of visual processing [27]. In vision, objects tend to occupy spatially compact regions of the visual field. Long range connections are therefore inappropriate, since they could lead to the synchronisation of neural oscillators that represent features of different objects. However, spatial compactness is a less useful cue in audition; indeed, the auditory system *must* combine information from different spatial (frequency) regions in order to build a perceptual description of wideband signals, such as speech. It is doubtful, therefore, whether the lateral connections in Wang's model are able to facilitate the grouping of acoustic components that are widely separated in frequency.

Another feature of two of the models reviewed in Section 2 is that they employ global inhibitory mechanisms. The function of the H-cell in Malsburg & Schneider's model and the global inhibitor in Wang's model is the same; to desynchronise groups of oscillators that have a weak mutual coupling. In other words, global inhibition introduces an element of competition into the organisation of an oscillator network, so that only relatively strong coupling leads to temporal synchronisation. The Brown & Cooke model does not address the desynchronisation of weakly coupled groups. In the model of Liu *et al*., desynchronisation is promoted by inhibitory links between next-nearest neighbours in the B-centre.

The global inhibitor approach has two advantages. First, the activity of the global inhibitor is a good indicator of the state of synchronisation in an oscillator network. Second, the behaviour of the H-cell/global inhibitor may correspond to that of a neural group in the thalamus [8]; the approach therefore has some physiological justification.

3.3. The representation of time

None of the models reviewed in Section 2 adequately addresses the representation of time in the auditory system. The Liu *et al*. and Malsburg & Schneider models assume that the input is a time series of spectral patterns, but they provide no means of comparing the properties of patterns which arrive at different time instants. The coupling strengths between oscillators in the Brown & Cooke model are maintained for some time after stimulation has ceased, and therefore act as a form of short-term memory. This allows their model to simulate the build-up of streaming over time (see below) but, again, does not allow auditory activity to be compared at different time instants.

In Wang's model, time and frequency are represented on the axes of a two-dimensional grid of neural oscillators. This 'neural spectrogram' approach certainly has intuitive appeal; it is well known that the auditory system maintains a tonotopic (frequency) axis, so it is tempting to suppose that it also maintains a time axis. However, there is no physiological evidence for systematic arrangements of neurons that are capable of generating the long delays needed for auditory grouping (in excess of 200 ms). Indeed, such systems of delay lines may not be theoretically possible. In order to obtain the temporal resolution required for auditory organisation, a delay-line scheme of the form proposed by Wang would involve several hundred synapses. Transmission in simple chains of neurons of this length is unlikely to be precisely timed, due to jitter in conduction time and synaptic processes [1].

Another reason for hypothesising an auditory time axis is that it allows acoustic stimuli to be treated as two-dimensional spatial patterns; this is advantageous, since the architecture of Wang's model is closely related to that of his earlier models of visual processing [27]. However, caution is needed when drawing analogies between the auditory and visual modalities. In vision, time and space are two separate characteristic of a stimulus arriving at the retina, whereas in audition temporal and spatial (frequency) dimensions are intrinsically linked. Hence, in the visual pathway the temporal and spatial characteristics of a stimulus may be completely separable, whereas in the auditory pathway they may not.

A final point related to the representation of time is that only the Brown & Cooke model takes account of the *time course* of auditory organisation. Indeed, Wang's oscillator network requires just n cycles of oscillations to organise an input consisting of n simultaneously presented patterns, so long as the parameter ε in equation (4b) is sufficiently small. While this is clearly advantageous for real-time signal processing, it conflicts with auditory psychophysics. For example, it may take ten seconds or so for a sequence of alternating tones to segregate into two perceptual streams [5,6].

3.4. Exclusive allocation vs. joint allocation

It is usually the case that the auditory system assigns each acoustic feature to a single stream; the so-called *principle of exclusive allocation*. However, there are many exceptions to this rule, such as in the 'duplex' perception of speech and the perception of harmonic complexes with a mistuned component [5]. Such violations of the principle of exclusive allocation have a clear implication for neural oscillator theories; neurons coding shared acoustic features should be permitted to synchronise with two (otherwise desynchronised) groups of oscillations.

Of the models described in Section 2, only Wang's scheme meets this criterion. The relaxation oscillators employed in his model can exhibit aperiodic firing activity, and may therefore synchronise to each of two patterns in the input. In contrast, the chaotic oscillators used in the Brown & Cooke model may only synchronise with a single group. Since Liu *et al*. only address the recognition of single vowels, it is not clear whether their model allows the same formant to be allocated to two sources, as might occur in the 'duplex' perception of speech.

3.5. Discrete vs. continuous input representations

The models described in Section 2 employ different input representations. The Brown & Cooke scheme includes a realistic simulation of peripheral auditory processing, whereas the model of Liu *et al*. employs a LPC front-end. The input to

the remaining models consists of simulated time-frequency patterns. Further, the majority of the models require binary, rather than continuously valued, inputs. For example, a pair of relaxation oscillators in the Wang and Liu *et al*. models must receive almost identical input values in order to approach the same limit cycle in the phase plane. Hence, Wang's oscillators receive input values of -0.02 (no stimulation) or 0.2 (stimulation), and Liu *et al*. normalise their LPC coefficients so that they are either zero or in the range [2.5, 3.5].

It is generally assumed that place-coded maps of features in the auditory system provide probabilistic information, such that the firing rate of a cell in the map indicates the likelihood that a particular feature is present [22]. However, if the dynamics of neural oscillator models accurately reflect the mechanisms of feature binding in the auditory system, they imply that the representation of features in maps is discrete rather than probabilistic. In other words, the presence of a feature would be indicated if the firing activity in the appropriate location of a map exceeded a threshold value. A further implication is that sound intensity would then need to be coded and bound in a separate map. Recent physiological studies of the cortex suggest that this is a possibility [24].

3.6. The number of concurrent auditory streams

What limitations does the neural oscillator theory place on the number of concurrent auditory streams? It has been argued that listeners maintain multiple auditory streams, even though they are not the subject of conscious attention [5]. However, the maximum number of concurrent streams is probably no more than three. Hence, neural oscillator architectures should permit three synchronised blocks of oscillations to exist concurrently.

It has been noted that certain classes of laterally-coupled neural oscillator networks are only able to represent a very small number of simultaneous patterns [7]. However, at least two of the approaches discussed in Section 2 do not suffer from this problem. A detailed mathematical analysis of Wang's oscillator network has shown that it is able to represent *n* concurrent patterns in the input as *n* blocks of synchronised oscillations [26]. The sine circle maps employed in Brown & Cooke's model exhibit chaotic oscillations, which allow a large number of groups to be represented; the number of possible groups is limited only by the resolution of the phase variable θ in equation (6). It is unclear from the simulations presented by Malsburg & Schneider and Liu *et al*. whether their models are able to represent more than two concurrent patterns.

3.7. The default state of auditory organisation

With the exception of the Brown & Cooke scheme, the models described in Section 2 assume that the default state of auditory organisation is segregation; in other words, features remain segregated unless there is evidence for grouping them. However, it seems likely that the auditory system fuses acoustic components by default. For example, listeners initially perceive sequences of tones that alternate between widely separated frequency regions as a single source [5,6].

If fusion is the default state of auditory organisation, the neural oscillator theory predicts that local field potentials recorded from the higher auditory system will be synchronised at the onset of a stimulus, and then become desynchronised as evidence for perceptual segregation is accumulated.

4. CHALLENGES FOR OSCILLATOR THEORIES

Oscillator models for auditory organisation have received far less attention than their visual counterparts. Consequently, those attempts reported above should be regarded as initial forays. Many issues peculiar to the auditory domain are only now being discovered. This section discusses some of challenges which future models of binding, including neural oscillator models, will be forced to address.

4.1. Neural oscillations, memory and learning

The same cortical circuitry supports memory and information processing, so the mechanisms of computation and the mechanisms of memory must be closely related. Indeed, memory should be considered an integral part of the binding problem: not only must we bind together a combination of features at one moment, but we must also store that combination of features for future reference [9]. Further, there appears to be an intimate connection between oscillations, synchronisation and memory. Synchronous bursts converging on a postsynaptic cell produce large depolarisations that are optimal for activating NDMA receptors, which in turn lead to long-term potentiation. Hence, neural plasticity may require temporal synchronisation of synaptic inputs [23].

If neural oscillations are involved in long-term memory and learning, then they must in some way account for perceptual invariance: the problem that many different sensory experiences must be mapped onto the same internal representation. Malsburg's correlation theory [19] proposes a solution to perceptual invariance in terms of a mechanism of associative memory, in which independent sets of correlations lead to independent associative mappings. Based on this idea, Wang *et al*. [25] have described an associative memory which consists of coupled oscillators. Their network is able to perform segregation of simultaneously presented patterns which have previously been stored in the network. Segregation in this model is therefore schema-driven [5]; primitive (data-driven) mechanisms are not included.

Regarding short-term memory (STM), Horn & Usher [15] describe an oscillatory model which accounts for the limited (7+/-2) capacity of STM in terms of competition between oscillations. This is achieved through a short-term potentiation mechanism, which lowers the threshold of oscillating cells after stimulation has ceased, causing oscillations to persist without input. Lisman & Idiart [17] elaborate a compatible physiologically-motivated scheme based on the storage of memories in different high frequency (40 Hz) subcycles of a low frequency (5-12 Hz) oscillation, resulting in the 7+/-2 constraint. It should be noted that such capacity limitations are not present in the Bauer & Martienssen model [4] on which the Brown & Cooke scheme [6] is based; whilst this has engineering advantages, their model is not in accord with human performance.

4.2. Binding of features in multiple maps

It was noted in the introduction that the binding problem may operate both within processing nuclei and between such regions. In the former, it may reflect the grouping of elements within the same auditory map (e.g., binding tonotopic channels that possess similar rates of envelope modulation). In the latter case, it might refer to the binding of properties processed separated (e.g., pitch and loudness). The formulation of the binding problem usually implies the latter interpretation, but, oddly, most modelling studies – including those reviewed in Section 2 – only address within-region binding.

Damasio [9] describes an architecture for binding information from multiple brain regions. He hypothesises a hierarchy of neural groups called convergence zones, which trigger and synchronise neural activity patterns in lower centres. Convergence zones store binding codes, which describe the combination of features that describe entities. Patterns of neural activity that correspond to properties of entities are recorded in the same neural groups in which they occur during perception. During perception and recall, convergence zones achieve synchronisation of these features through recurrent feedback interactions. Hence, processing is distributed and does not occur in a single direction, nor does it require integration in a single area; it involves phase-locking amongst neural groups in multiple regions (see also [2]).

4.3. Evidence against the neural oscillator theory

Aspects of the neural oscillator theory are controversial. In particular, it has been suggested that neural oscillations are induced by anaesthesia, and that oscillations therefore reflect state of sleep rather than feature binding [14]. There is some justification for this view, since many studies reporting cortical oscillations have been conducted on animals using anaesthetics that are known to induce rhythmic neural firing activity. Equally, however, studies have reported coherent neural oscillations in awake animals [23]. It seems unlikely, therefore, that oscillations are simply an epiphenomenon.

It should be noted that evidence for coherent oscillations in the auditory system currently comes only from imaging techniques such as the EEG (e.g., [13,16]). Such noninvasive studies are perhaps more open to interpretation than recordings of local field potentials from macroelectrodes. The neural oscillator theory of auditory feature binding would therefore be considerably strengthened if evidence from invasive recording techniques was available.

4.4. Oscillations and synchrony

It should be stressed that von der Malsburg's correlation theory [19] makes a general statement about brain function which only implies the synchronisation of neural firings; it does not require neural oscillations *per se*. Indeed, the dynamical system considered in Section 1.2 suggests that isolated neurons can exhibit oscillatory activity without being synchronised to other cells. Similarly, the firing activity of different cells can be synchronised without exhibiting oscillations. In other words, although neural oscillations and synchronisation often occur together, they do not necessarily depend on one another.

For example, it has been demonstrated that synchronous transmission in chains of neurons with diverging and converging links (so-called 'synfire' chains) can lead to oscillations [2]. In such networks, oscillations do not occur because of periodic activation of the same cells; rather, they are due to the interaction of positive (excitatory) feedback and negative (inhibitory) feedback. Hence, oscillations may occur as a consequence of synchronisation, rather than *vice versa*.

ACKNOWLEDGEMENTS

GJB and MPC are supported by EPSRC grants GR/H53174 and GR/K18962. EM is supported by research fellowship CHRX-CT93-0098 from the EU HCM programme.

REFERENCES

[1] Abeles, M. (1991) *Corticonics: Neural circuits of the Cerebral Cortex*. Cambridge University Press.

[2] Abeles, M., Prut, Y., Bergman, H. & Vaadia, E. (1994) Synchronisation in neuronal transmission and its importance for information processing. *Progress in Brain Research*, **102**, 395-404.

[3] Barlow, H. (1972) Single units and sensation: A neuron doctrine for perceptual psychology? *Perception*, **1**, 371-94.

[4] Bauer, M. & Martienssen, W. (1991) Coupled circle maps as a tool to model synchronisation in neural networks. *Network*, **2**, 345-51.

[5] Bregman, A. (1990) *Auditory scene analysis*. MIT Press.

[6] Brown, G. & Cooke, M. (1995) Temporal synchronisation in a neural oscillator model of primitive auditory stream segregation. *Proceedings of the IJCAI Workshop on Computational Auditory Scene Analysis*, Montreal, 40-47.

[7] Cairns, D., Baddeley, R. & Smith, L. (1993) Constraints on synchronising oscillator networks. *Neural Computation*, **5**, 260-66.

[8] Crick, F. (1984) Function of the thalamic reticular complex: The searchlight hypothesis. *Proceedings of the National Academy of the Sciences USA*, **81**, 4586-90.

[9] Damasio, A. (1989) The brain binds entities and events by multiregional activation from convergence zones. *Neural Computation*, **1**, 123-132.

[10] Eckhorn, R., Bauer, R., Jordan, W., Brosch, M., Kruse, W., Munk, M. & Reitboecj, H. (1988) Coherent oscillations: A mechanism of feature linking in the visual cortex. *Biological Cybernetics*, **60**, 121-130.

[11] FitzHugh, R. (1961) Impulses and physiological states in theoretical models of nerve membrane. *Biophysical Journal*, **1**, 445-66.

[12] Frisina, R., Smith, R. & Chamberlain, S. (1990) Encoding of amplitude modulation in the gerbil cochlear nucleus: A hierarchy of enhancement. *Hearing Research*, **44**, 99-122.

[13] Galambos, R., Makeig, S. & Talmachoff, P. (1981) A 40 Hz auditory potential recorded from the human scalp. *Proceedings of the National Academy of the Sciences USA*, **78**, 2643-47.

[14] Horikawa, J., Tanahashi, A. & Suga, N. (1994) After-discharges in the auditory cortex of the moustached bat - no oscillatory discharges for binding auditory information. *Hearing Research*, **76**, 45-52.

[15] Horn, D. & Usher, M. (1991) Parallel activation of memories in an oscillatory neural network. *Neural Computation*, **3**, 31-43.

[16] Joliot, M., Ribary, U. & Llinas, R. (1994) Human oscillatory brain activity near to 40 Hz coexists with cognitive temporal binding. *Proceedings of the National Academy of the Sciences of the USA*, **91**, 11748-51.

[17] Lisman, J. & Idiart, M. (1995) Storage of 7+/-2 short-term memories in oscillatory subcycles. *Science*, **267**, 1512-15.

[18] Liu, F., Yamaguchi, Y. & Shimizu, H. (1994) Flexible vowel recognition by the generation of dynamic coherence in oscillator neural networks: speaker-independent vowel recognition. *Biological Cybernetics*, **7**, 105-114.

[19] von der Malsburg, C. (1981) *The correlation theory of brain function*. Internal report 81-2, Max Planck Institute for Biophysical Chemistry, Göttingen, FRG.

[20] von der Malsburg, C. & Schneider, W. (1986) A neural cocktail-party processor. *Biological Cybernetics*, **54**, 29-40.

[21] McGurk, H. & McDonald, J. (1976) Hearing lips and seeing voices. *Nature*, **264**, 746-48.

[22] Moore, D. (1987) Physiology of higher auditory system. *British Medical Bulletin*, **43**, 856-70.

[23] Singer, W. (1993) Synchronisation of cortical activity and its putative role in information processing and learning. *Annual Review of Physiology*, **55**, 349-74.

[24] Sutter, M. & Schreiner, C. (1995) Topography of intensity tuning in cat primary auditory cortex - single-neuron versus multiple-neuron recordings. *Journal of Neurophysiology*, **73**, 190-204.

[25] Wang, D., Buhmann, J. & von der Malsburg, C. (1990) Pattern segmentation in associative memory. *Neural Computation*, **2**, 94-106.

[26] Wang, D. (1996) Primitive auditory segmentation based on oscillatory correlation. *Cognitive Science*, in press.

[27] Wang, D. (1995) Image segmentation based on oscillatory correlation. *Proceedings of the World Congress on Neural Networks*, in press.