ABSTRACT

Title of dissertation: PROCESSING INFORMATION ON INTERMEDIATE TIMESCALES WITHIN RECURRENT NEURAL NETWORKS

Oliver L. C. Rourke, Doctor of Philosophy, 2016

Dissertation directed by: Professor Dan A. Butts
Department of Biology

The cerebral cortex has remarkable computational abilities; it is able to solve problems which remain beyond the most advanced man-made systems. The complexity arises due to the structure of the neural network which controls how the neurons interact. One surprising fact about this network is the dominance of 'recurrent' and 'feedback' connections. For example, only 5-10% of connections into the earliest stage of visual processing are 'feedforward', in that they carry information from the eyes (via the Lateral Geniculate Nucleus). One possible reason for these connections is that they allow for information to be preserved within the network; the underlying 'causes' of sensory stimuli usually persist for much longer than the time scales of neural processing, and so understanding them requires continued aggregation of information within the sensory cortices. In this dissertation, I investigate several models of such sensory processing via recurrent connections. I introduce the transient attractor network, which depends on recurrent plastic connectivity, and demonstrate in simulations how it might be involved in the processes of short term

memory, signal de-noising, and temporal coherence analysis. I then show how a certain recurrent network structure might allow for transient associative learning to occur on the timescales of seconds using presynaptic facilitation. Finally, I consider how auditory scene analysis might occur through 'gamma partitioning'. This process uses recurrent excitatory and inhibitory connections to preserve information within the neural network about its recent state, allowing for the separation of auditory sources into different perceptual cycles.

PROCESSING INFORMATION ON INTERMEDIATE TIMESCALES WITHIN RECURRENT NEURAL NETWORKS

by

Oliver L. C. Rourke

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2016

Advisory Committee:
Professor Dan Butts, Chair/Advisor
Professor Radu Balan
Professor Wojciech Czaja
Professor David Jacobs
Professor Quentin Gaudry

Table of Contents

List of Figures

List of Abbreviations

Hz	Hertz
LEGION	Local Excitatory Global Inhibitory Oscillator Network
LTD	Long Term Depression
LTP	Long Term Potentiation
PC	Principal Component (from PCA)
PCA	Principal Component Analysis
PFC	Prefrontal Cortex
PNG	Polysynchronous Neuronal Group
PPV	Positive Predictive Value
s	Second
STM	Short Term Memory
STP	Short Term Potentiation
TPR	True Positive Rate

A note on shorthand for synaptic strengths.
S = Stimulus, E = Excitatory, I = Inhibitory.
Numbers indicate layer (where appropriate): E1 = first layer of excitatory.
X is a wildcard for a number.
W_{AXBX} = Weight from type A in any layer to type B in the same layer.

Chapter 1: Introduction

1.1 Basic Anatomy and Physiology of a Neuron

Neurons are a type of cell responsible for receiving, processing, and transmitting information. Although there are various types of neurons, they share three main anatomical features: the dendrites, the cell body, and the axon (Figure 1.1).

Inputs to the cell body come from the neuron's dendrites. These inputs change the neuron's membrane potential, causing either an increase (excitatory inputs) or a decrease (inhibitory inputs). Should the membrane potential be raised to a sufficiently high level, the neuron generates an action potential, also known as a spike. Although action potentials do vary somewhat in their duration, amplitude, and shape (Dayan and Abbott, 2001), they are typically treated as identical events in neural encoding. This action potential then propagates along the axon, at the end of which are located junctions with other cells. These junctions are known as synapses, and cause a postsynaptic current (PSC) into the postsynaptic cell whenever there is a spike in the presynaptic cell. Although the presynaptic spikes are similar, the magnitude of the PSC varies in accordance with some synaptic strength. The factors that influence this synaptic strength are summarized below (Section 1.3).

All neurons may be classified as either excitatory or inhibitory depending on their effect on any postsynaptic neurons. Spikes from excitatory cells always induce a positive PSC, whereas those from inhibitory cells induce a negative PSC. This di-

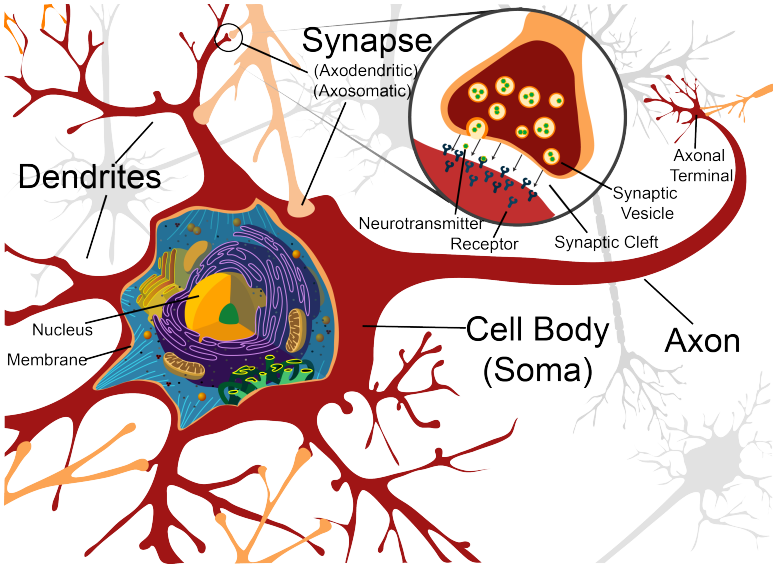

Figure 1.1: **A diagram of a neuron.** A simple diagram showing cell body, dendrites, and axon. *Inset:* Details of a chemical synapse. Modified from 'Complete neuron cell diagram.svg' by LadyOfHats which has been released into the public domain.

vision of excitatory vs. inhibitory is decided by cell type; a single neuron can not contain both excitatory and inhibitory synapses, and cells can not transform from excitatory to inhibitory.

1.2 Neural Processing

The processes underlying neural behavior may be quantified; indeed, this is necessary if we are to construct computational simulations of networks of neurons. In this section I briefly outline two models that I use in later chapters.

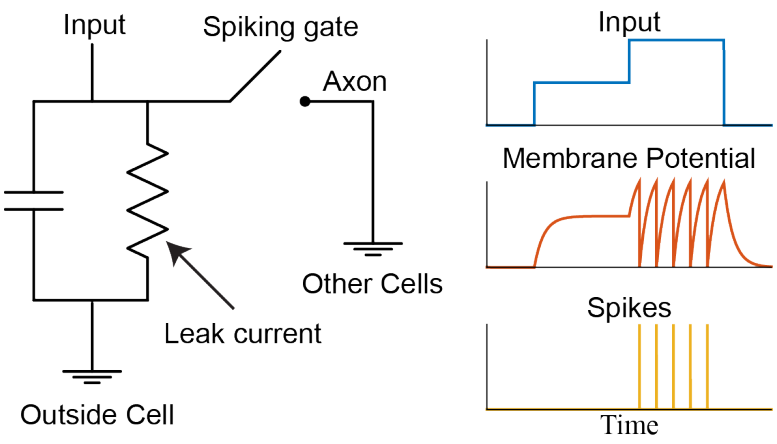

Figure 1.2: **LIF neuron as a circuit** *Left:* Circuit diagram. *Right:* Neuron behavior.

1.2.1 Leaky Integrate-and-Fire model

In the Leaky Integrate-and-Fire (LIF) model, the neuron's state is determined by its membrane potential, $V_j(t)$. This is governed by the derivative form of the capacitor equation with a leak current and both excitatory and inhibitory synaptic input currents,

$$C\frac{dV_j}{dt} = I_j^{Leak}(t) + I_j^E(t) + I_j^I(t). \tag{1.1}$$

. This is equivalent to a simple circuit (Figure 1.2). The leak current can then be derived from Ohm's Law

$$I_j^{Leak}(t) = g_m[V_{rest} - V_j(t)], \tag{1.2}$$

where g_m is the cell membrane conductance, and V_{rest} is the resting potential. The synaptic input currents are calculated by summing the currents from each individual

synapse:

$$I_j^E(t) = \sum_{i=1}^{N_E} g_{ij}^E(t)[E_{rev}^E - V_j(t)] \tag{1.3}$$

$$I_j^I(t) = \sum_{i=1}^{N_E} g_{ij}^I(t)[E_{rev}^I - V_j(t)]. \tag{1.4}$$

Terms of the form E_{rev}^* represent the reversal potentials for the excitatory and inhibitory currents. We note that $E_{rev}^I < V < E_{rev}^E$ in a cell which is near its spiking threshold, meaning an increase in excitatory conductance will lead to a positive current (and vice versa for inhibitory). At rest, the excitatory and inhibitory currents are balanced. Presynaptic spikes change the conductances of the channels, which in turn leads to a shift in the cell membrane potential. In this dissertation, the term synaptic weight (notated as W_{ij}^E and W_{ij}^I) will be used synonymously with the synaptic conductance (g_{ij}^E and g_{ij}^E).

The LIF model does not explicitly contain a description of spike generation. Instead, a firing threshold is set. Should the membrane potential reach that threshold, a binary spike is produced, and the neuron's membrane potential is reset to its rest value:

$$t_{spk} : V_j(t_{spk}) = \theta \tag{1.5}$$

$$\lim_{t \to t_{spk}^+} V_j(t) = V_{rest}. \tag{1.6}$$

Alternatively, in order to approximate inactivated Na^+ channels after each spike,

we may include an absolute refractory period, t_{ref}:

$$I_j^E(t) = \begin{cases} 0 & t_{spk} < t < t_{tpk} + t_{ref} \\ I_j^E(t) & \text{otherwise} \end{cases} \tag{1.7}$$

(Brody and Hopfield, 2003; Miconi and Vanrullen, 2010; Mongillo et al., 2008).

Neural behavior can appear to be highly random; there is a wide variability in firing rates of neurons within sensory cortices between trials with the same sensory stimulus (Fourcaud and Brunel, 2002; Masquelier, 2013; Movshon, 2000). This variation comes from a variety of external and intrinsic properties; a large amount of effort has gone into teasing these two sources apart. In neural models, however, it is common to combine these factors into one additive noise term,

$$C\frac{dV_j(t)}{dt} = -RV_j(t) + I_{Total,j}(t) + \alpha_N dW_t, \tag{1.8}$$

where dW_t is the time derivative of Brownian motion ($N(0, \sqrt{\Delta t})$). This additive noise is of particular interest when subpopulations of neurons receive the same excitatory and inhibitory currents, but where variability within the subpopulations is important.

1.2.2 Firing Rate Models

For large networks of neurons, individually modeling each neuron's membrane potential can be computationally expensive; it is often easier to consider groups of neurons with similar properties and track the averaged firing rates (Ermentrout and Terman, 2010). The firing rate may also be thought of as each individual neurons

probability of spiking; the probability of spiking over time is assumed to be constant for any given stimulus even if individual spike times are highly variable. Such firing rate models are the basis for most artificial neural networks used in machine learning. We now construct a typical firing rate model for use in simulating biological networks.

The dynamics of the firing rate model are highly similar to those of the LIF model described above. The synaptic currents and membrane potential are calculated using equations 1.1 to 1.4. The firing rate is then calculated as a nonlinear function applied to a hidden variable,

$$y_j(t) = F(V_j(t)). \tag{1.9}$$

In this dissertation, I use a saturating, rectified function $F(x) = \max[0, 1-\exp(-a(x-b))]$ (Figure 1.3). Other popular options include the logistic, rectified linear, and softplus functions.

1.3 Synaptic Plasticity

The function of networks of neurons is dictated by the strengths of the synapses which connect the neurons to one another. These strengths are constantly changing through processes known as synaptic plasticity. These changes in synaptic strengths can therefore have a significant influence on the function of these networks, and is a point to which we return several times throughout this dissertation. This section, therefore, gives a brief biological and mathematical overview of synaptic plasticity.

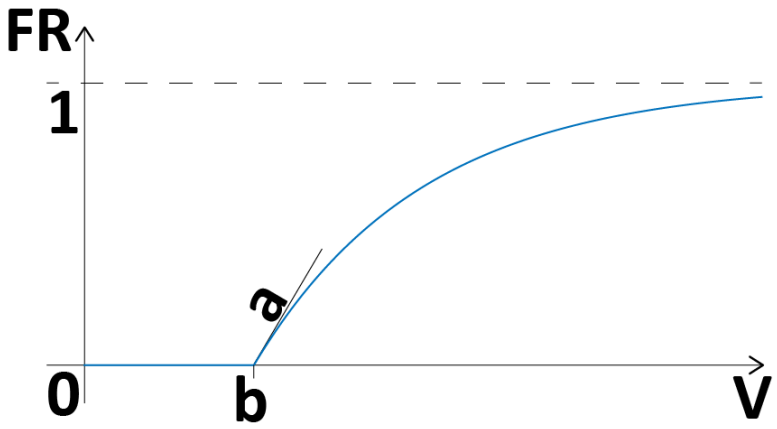

Figure 1.3: **Firing rate nonlinearity.** Saturating, rectified function which may be used in firing rate models.

We divide these synaptic plasticity processes into two categories, long-term and short-term, depending on the time scales for which the effects persist. Long-term plasticity plays a key role in deciding how networks of neurons are typically connected together, whereas short-term plasticity may play a significant role in affecting how networks of neurons process incoming information.

1.3.1 Long-Term Plasticity

Long-term plasticity encompasses mechanisms that lead to changes in synaptic connectivity that persist for long amounts of time (at least several minutes). Two forms of long-term plasticity are long-term potentiation (LTP) and long-term depression (LTD), which respectively strengthen and weaken the synapses. The mechanisms underlying LTP and LTD are not completely understood, and these processes may vary significantly between different cortical regions. For example, experiments by

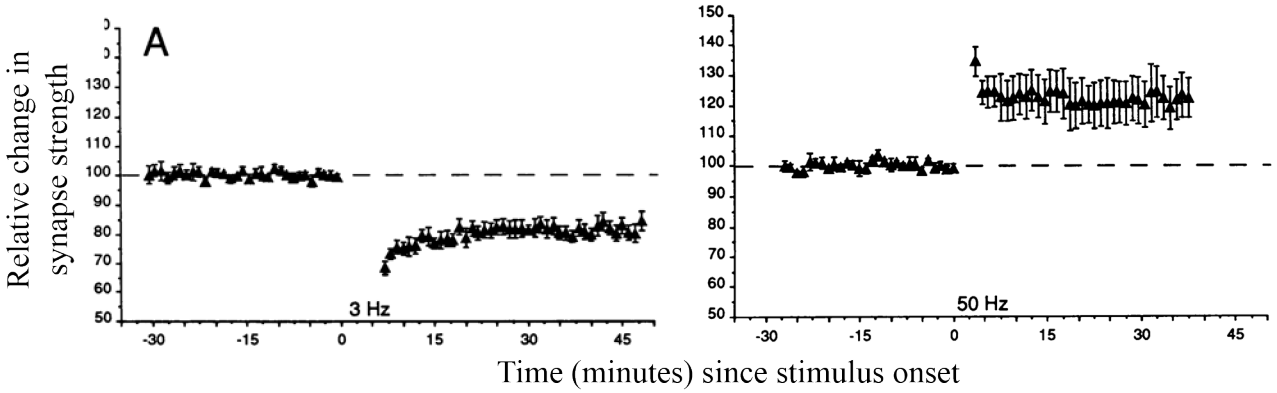

Figure 1.4: **Hebbian LTP/LTD.** Experimental recordings of Hebbian plasticity (Dudek and Bear, 1992). When paired with postsynaptic activity, low rates of presynaptic activity lead to LTD (*left*) while high rates lead to LTP (*right*).

Dudek and Bear (1992) in the hippocampus revealed LTP in the presence of high presynaptic firing rates, and LTD in the presence of low presynaptic firing rates (Figure 1.4). These long-term plasticity mechanisms, when combined with the history of the network's activation, play a significant role in establishing the network connectivity (the synaptic strengths between the various neurons).

One possible explanation for the adaptation of synapses is provided by Hebbian theory (Hebb, 1949). The theory suggests that changes to plasticity follow an associative learning rule, as summarized by the maxim

"Cells that fire together, wire together"

(Shatz, 1992). This is a form of correlation-based learning, similar to classical ('Pavlovian') conditioning. In trying to decode causation behind stimuli, the network records correlation. This associative plasticity is believed to play a role in

the formation of memories and learning within neural networks, and is a topic to which we will return when discussing the Hopfield Network (Section 1.4.2.1), and information storage within neural networks (Section 1.5).

1.3.2 Short-Term Synaptic Plasticity

Synaptic strengths are also known to undergo temporary modifications due to neural activity; these effects typically decay within seconds. These processes may be classified as either facilitation or depression depending on whether they increase or decrease synaptic strengths. Multiple different processes have often been found to function at individual synapses, and these processes may function to varying extents in different synapses (Hennig, 2013; Regehr, 2012).

1.3.2.1 Presynaptic Facilitation

Several mechanisms are known to increase the effective transmission of signals following periods of elevated presynaptic firing. Decay time is often used to categorize such effects into facilitation (decays over tens of milliseconds), augmentation (decays over seconds), and post-tetanic potentiation (decays over tens of seconds to minutes) (Abbott and Regehr, 2004; Purves et al., 2012). These processes have all been linked to the build up of Ca^{2+} in the presynaptic terminal caused by the large amount of presynaptic activity. In the neural network modeling literature, these processes are typically combined together under the name of 'facilitation', with an

assumed decay time of a few seconds (Barak and Tsodyks, 2007; Itskov et al., 2011; Mejias and Torres, 2009; Mongillo et al., 2008). Such a process can be modeled through the inclusion of an additional presynaptic facilitation variable, $u_i(t)$, which has a multiplicative effect on the synaptic strength

$$W'_{ij}(t) = u_i(t)W_{ij}. \tag{1.10}$$

The facilitation variable is governed by

$$\frac{du_i(t)}{dt} = \frac{1}{\tau_{u_+}} \left(u_{max} - u_i(t)\right) y_i(t)^\alpha - \frac{1}{\tau_{u_-}} \left(u_i(t) - 1\right) \tag{1.11}$$

(Tsodyks et al., 1998). Thus in the presence of high presynaptic firing rates, $y_i(t)$, the facilitation variable increases towards an upper limit u_{max}, and decays towards 1 in the absence of presynaptic firing.

1.3.2.2 Presynaptic Depression

Another mechanism exists which reduces the effective synaptic strength following significant presynaptic activity; this is known as synaptic depression. This occurs due to a depletion of resources in the presynaptic terminal, and decays over around 100ms. This process is modeled in a similar manner to the enhancement above by incorporating a presynaptic depression variable $x_i(t)$,

$$W''_{ij}(t) = x_i(t)W'_{ij}(t), \tag{1.12}$$

which is governed by

$$\frac{dx_i(t)}{dt} = \frac{1}{\tau_{x_+}} \left(1 - x_i(t)\right) y_i(t)^\alpha - \frac{1}{\tau_{x_-}} x_i(t)u_i(t)y_i(t) \tag{1.13}$$

(Tsodyks et al., 1998). In the presence of presynaptic activity, the depression variable is driven towards some lower limit (0), and rebounds towards 1 in the absence of firing.

1.3.2.3 Transient Associative Plasticity

The above short-term mechanisms all depend solely on the pre- or postsynaptic firing rates, not the combination of the two. It has also been conjectured that a short-term associative effect might exist (Brenowitz and Regehr, 2005), modifying synaptic strengths according to both pre- and postsynaptic firing rates. This process has been associated with several tasks including short-term memory (Sandberg et al., 2003; Szatmáry and Izhikevich, 2010), auditory streaming (Von der Malsburg and Schneider, 1986), visual object separation (Becker and Plumbley, 1996), corrective prediction in behavioral tasks (Schultz and Dickinson, 2000), and in modification to persistent activity (Brunel, 2003). Additionally, experiments in vivo (Shamma et al., 2013; Sugase-Miyamoto et al., 2008) have suggested that network connectivity does have some short time-scale associative properties.

One possible process for this is 'Short-Term Potentiation'; sometimes, following activity that would normally cause LTP, the synapse is only momentarily strengthened before returning to its original state. The time scales involved for strengthening are nevertheless on the order of several minutes (Erickson et al., 2011; Hennig, 2013; Malenka, 1991), beyond the 'intermediate time-scales' considered in this disserta-

tion.

1.4 From Neurons to Networks of Neurons

In the body of this dissertation, the majority of the work is done using computer simulations of large populations of neurons and synapses. The rules which individually govern each part have been discussed above in Sections 1.2 and 1.3. Understanding their collective action, however, requires understanding how the network as a single entity behaves. We briefly discuss a few such issues in this section. We start by considering what forms of connectivity are typically present in networks of neurons, and then describe how their collective behavior might be understood by examining the attractors.

1.4.1 Structures in Networks of Neurons

The manner in which stimuli are processed in the sensory cortices is largely decided by the connectivity within the networks of neurons. Understanding the general connectivity pattern is therefore of great importance when constructing models of sensory processing. In this section, we summarize a few key facts that are known about the structure of cortical networks of neurons, and then discuss how this structure might allow for various functional objectives to be achieved.

1.4.1.1 Hierarchical Processing in Networks of Neurons

One of the early significant insights into how neural networks process information was made by observing how neurons in the primary visual cortex of an anesthetized cat respond to visual stimuli. It was already known that inputs from the retina were passed to a region known as the Lateral Geniculate Nucleus (LGN). Cells in the LGN had been observed to respond according to a 'center-surround' receptive field (RF), with excitatory connections at the centre of the RF surrounded by inhibitory inputs (Figure 1.5, *top*). Hubel and Wiesel (1959) recorded that some cells in the primary visual cortex were sensitive to a more complex feature, specifically to an oriented line at a particular point in space (Figure 1.5, *middle*). They called such cells 'simple'. They went on to discover a second type of neuron, 'complex' cells, which would respond to a line with a certain orientation but anywhere within a larger region of space (Hubel and Wiesel, 1962) (Figure 1.5, *bottom*). They realized that this could be explained by the sequential combination of features; each 'simple' cell's responses may be calculated by summing the output from a collinear set of LGN cells, and likewise 'complex' cells' may be formed by combining sets of nearby 'simple' cells with the same orientation (Figure 1.5). This led to the idea of hierarchical processing (Wurtz, 2009), that is that the cortex forms may detect complicated properties of stimuli by combining a variety of simpler properties. This form of processing has been investigated in a variety of sensory systems, including the visual (Felleman and Van Essen, 1991) and auditory (Kaas et al., 1999) systems, and can extend to higher-level recognition tasks. For example, the output

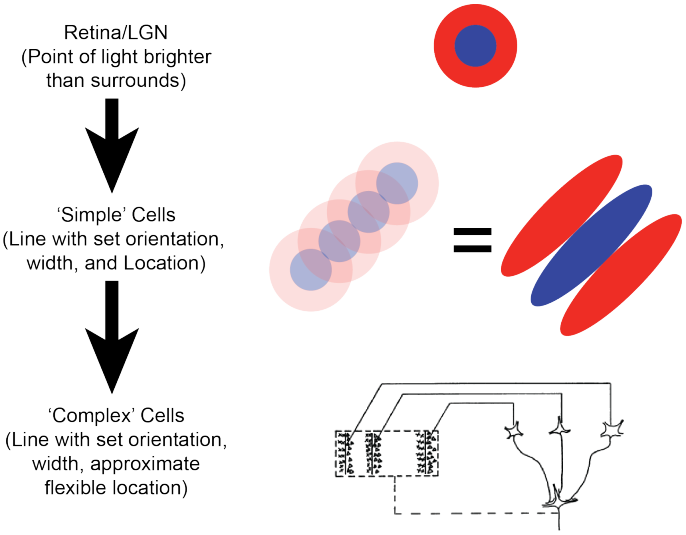

Figure 1.5: **Hierarchy in V1.** Successive combinations of simple features into more complicated features explains early receptive fields of neurons in early sensory processing.

from multiple simple or complex cells might be combined to detect the existence of a certain shape, output from multiple 'shape-sensitive' neurons could be combined to detect a certain object, and the output from multiple 'object-selective' could be combined to form abstract notions such as a group or identity.

This hierarchical organization of the sensory cortices has inspired many neural network models (Figure 1.6). Similar hierarchical neural network models have enjoyed great success in the fields of machine learning and machine learning.

1.4.1.2 Recurrent and Feedback Connections

The above hierarchical understanding of neural networks is enticing in its ability to explain how a neural structure might allow the network to perform higher order tasks such as object recognition. Investigations of the sensory cortices have, however,

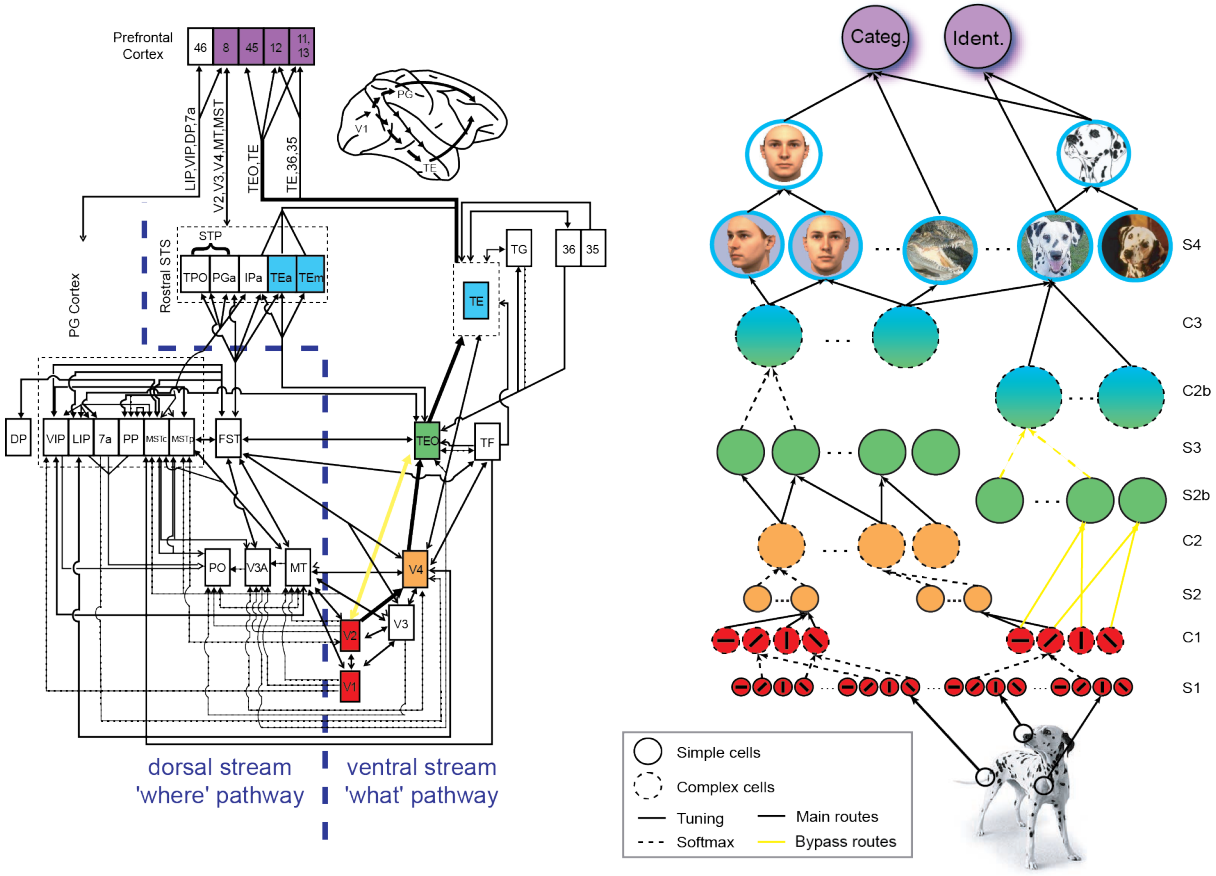

Figure 1.6: **Visual Processing Hierarchy.** A comparison of hierarchy in the visual cortex (*left*) and in an artificial neural network used for machine vision (*right*) (Serre et al., 2005).

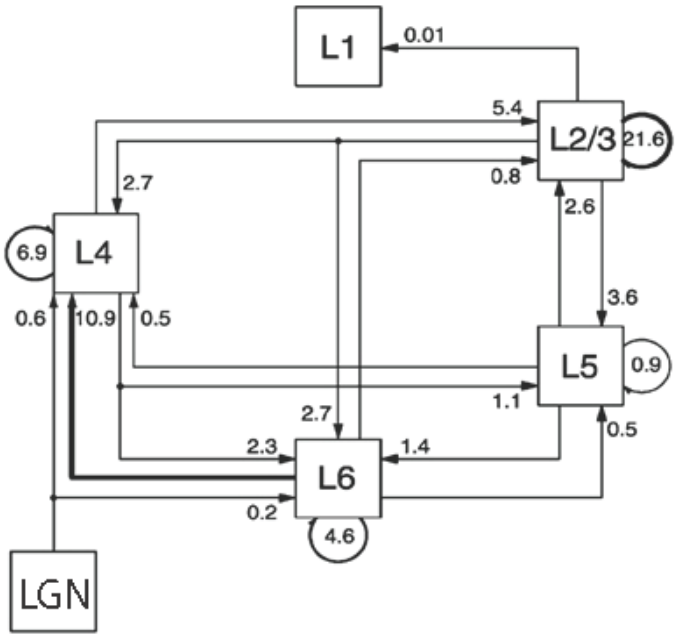

Figure 1.7: **Recurrent Network Structure.** Map of excitatory to excitatory synapses between different layers in the cat visual cortex. Numbers represent proportion of all excitatory to excitatory synapses in the region (about 30% could not be identified in this study). Only a minority of inputs come in from LGN, with no clear structural hierarchy present between the different layers.

suggested that this explanation of the network structure is, at best, incomplete. For example, only 5-10 % of the inputs into the monkey primary visual cortex come from the eyes (via the LGN) (Peters et al., 1994), suggesting much of what we 'see' is highly dependent on recurrent processing with the visual cortex (results from another study included in Figure 1.7). The reasons behind this connectivity are not well understood, and many different suggestions abound (Olshausen and Field, 2004). Indeed, this is a topic to which we return many times in the body of this dissertation. In the next section, we analyze the attractors in neural networks as dynamical systems in order to better understand the effects that these connections may have on the behavior of networks of neurons, and what functional outcomes might be achieved.

1.4.2 Attractors in Neural Networks

Several questions about neural network revolve around their long term behavior. For example: "Will the network ever settle into some steady pattern?", "How long does this take?", and "What types of patterns might we expect to see?". Questions such as these may be answered by examining the nature and location of the dynamic attractors within the neural network (Ermentrout and Terman, 2010; Izhikevich, 2007).

We initially define the attractor set of a neural network to be the set of asymptotically stable states in the network. This stability means that all points that are sufficiently close to the attractor evolve towards it; such points are said to lie in the basin of attraction. This also means that attractors are self-sustaining; once a network has entered an attractor, it remains there until it is made to change by some external force. As an illustration of attractors within neural networks, we start by considering a well known recurrent artificial neural network known as the 'Hopfield Network'. We then discuss other variants of attractors within neural networks which prove relevant to the work in the body of my dissertation.

1.4.2.1 The Hopfield Network

The Hopfield network is a recurrent neural network first proposed in 1982 (Hopfield, 1982). Many different variants exist; here we review the model as it was initially presented.

All units (neurons) in the model are binary, either on (1) or off (0). A set of N patterns is selected, with each pattern corresponding to a subset of all units being active. Using this set of patterns, weights are then set between all units using

$$w_{ij} = \frac{1}{N} \sum_{n=1}^{N} (2p_i^n - 1)(2p_j^n - 1) \tag{1.14}$$

where p_i^n represents the state of the ith neuron in the nth pattern, and w_{ij} is the weight between neurons number i and j (Figure 1.8A). This rule is associative in nature, meaning that strong positive connections form between neurons which are typically co-active (and negative connections between those with anti-correlated states). It is therefore broadly similar to the long term plasticity observed in cortical networks of neurons (Section 1.3), although the presence of both positive and negative weights coming from the same neuron is not biologically faithful.

The behavior of the neural units is governed by

$$V_i = H \left[\sum_{i \neq j} w_{ij} V_j - \theta \right] \tag{1.15}$$

where V_i represents the state of the ith unit, H is the Heaviside step function, and θ is the firing threshold. This is a firing-rate type model with instantaneous synaptic integration and a binary activation function. In the original formulation, each node is randomly selected to have its state updated, although deterministic variants with discrete time steps do exist.

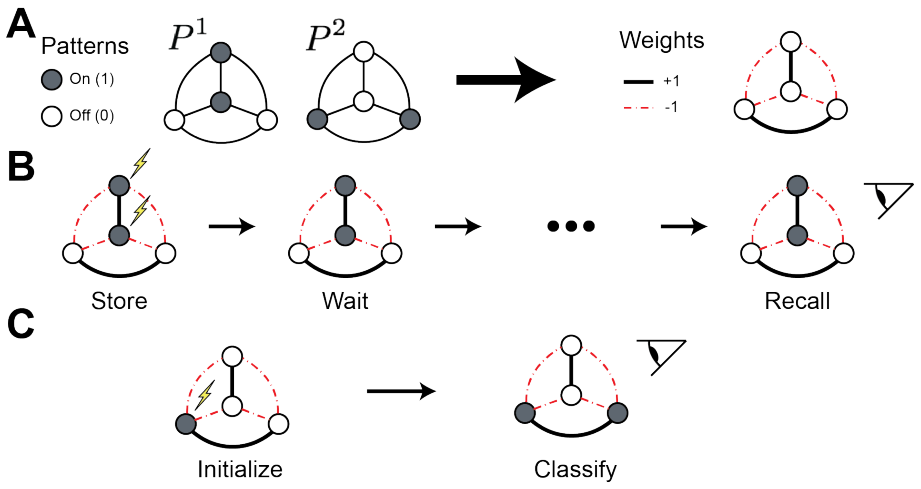

Figure 1.8: **The 'Hopfield Network'. A.** Associative training of Hopfield network. Two patterns P^1 and P^2) are used to determine recurrent weights. **B.** Information storage by activating an attractor. Activity is self-sustaining, allowing information to be later recalled by observing the state of the network. *C.* Different inputs may be classified by the basin of attraction in which they lie, which is often the pre-learned pattern with which they share the greatest similarity.

The network established above now contains multiple different attractors. Ideally, the set of attractors created matches with the set of patterns which was used to train the network. It is, however, possible that attractors exist which don't match with any of the patterns used, and that not all patterns have their own attractor (Orhan, 2014). We now consider two different functions that these attractors might achieve in the networks.

1.4.2.1.1 Temporary Information Storage

The attractors in the network are stable, and therefore self-sustaining. This allows for the temporary storage of information within the network, by activating one particular attractor (Figure 1.8B). This information persists unless interrupted, and may later be recalled by observing the state of the network. This form of informa-

tion storage does not require (or cause) some permanent changes to the networks structure. For this reason, attractors in neural network have often been used as a model for short-term memory. This is discussed further in Section 1.5 and Chapters 2 and 3.

1.4.2.1.2 Classifying Inputs

Novel incoming stimuli cause some pattern of activation within the network; from there, the attractor dynamics decide how the network behaves (Hopfield, 1982). In particular, the network will evolve towards some 'close' attractor, or more accurately to the attractor whose basin of attraction contained the initial activation pattern. This allows novel inputs to be mapped onto the set of attractors, ideally classifying novel inputs using the set of training patterns (Figure 1.8C). This behavior might allow neural networks to recognize patterns when they have been partially corrupted or distorted, or to understand novel patterns in terms of pre-learned classes.

The Hopfield network is a useful starting point when considering the roles attractors may play in neural networks. There are, however, several other properties of neural networks which may be attributed to the attractor dynamics which are not demonstrated in the classical Hopfield network.

1.4.2.2 Periodic Attractors and Cortical Oscillations

Oscillations in neural activity have been observed in many different regions of the brain and at a variety of frequencies. It has often been argued that these oscilla-

tions play multiple roles in cortical processing (Arnal and Giraud, 2012; Wang et al., 2010). For example, gamma waves (waves with frequency between 30 Hz and 100 Hz) have been associated with attention (Jensen et al., 2007), stimulus processing (Gray et al., 1989), short-term memory (Lundqvist et al., 2016), and motor control (Bragin et al., 1995). Moreover, gamma waves have been observed to synchronize over large areas of the cortex (Engel et al., 1990; Roskies, 1999), and with the degree of synchronicity being correlated with attention (Roelfsema et al., 1997). These gamma waves will play a crucial role in my model of stimulus processing in the auditory cortex (Chapter 4).

It is possible to better understand these oscillations by examining the network as a whole; in particular, these oscillations may be thought of as periodic attractors in the neural network (attractors which are not a single stable state, but a periodic cycle of states). Analyzing these attractors may provide deeper understanding into these oscillations. For example, Brunel and Hakim (1999) examined interconnected populations of excitatory and inhibitory populations for a variety of connection strengths. They discovered a Hopf bifurcation in the system when it transforms from having a fixed point attractor to a periodic attractor. These oscillations (the periodic attractor) exist because intervals of significant excitatory activity cause waves of dampening inhibitory activity. Properties of this periodic attractor match well with observed gamma waves. Other analysis has discussed lower frequency oscillations (Holcman and Tsodyks, 2006), and how different scales of oscillations might become coupled (Fontolan et al., 2013).

1.4.2.3 Slow Manifolds

The earlier approach to attractors only considered those which are asymptotically stable (all nearby points will approach the attractor). Alternatively, we may consider a broader family of attractors, those which are Lyapunov stable. Under this new definition, not all trajectories which start near the attractor need not tend towards it, however no trajectories may travel far away. In neural network literature, these attractors are often referred to as either line attractors, ring attractors, or plane attractors, depending on their topography (Druckmann and Chklovskii, 2010; MacNeil and Eliasmith, 2011; Sandberg et al., 2003; Seung, 1996). In the dynamics literature, this is known as the slow manifold, that is a subspace for a system of differential equations in which no eigenvalues are positive but there is at least one zero eigenvalue. In such systems, nearby trajectories approach the attractor (the manifold) and stay at one point on it (or move relatively slowly).

These types of attractors attractors have been proposed to allow for not just the storage of information, but also its aggregation (Goldman et al., 2007). The attractors essentially contain a continuum of stable points. This allows for the storage of continuous variables, and also allows for continual changes to be made. These slow manifolds have been hypothesized to allow for tracking eye location in the goldfish brain (Aksay et al., 2000), and physical location in the monkey brain (Wimmer et al., 2014).

1.4.2.4 Dynamic Field Theory (Bifurcation Theory)

The above discussion has considered attractors in a neural network to have predetermined locations and properties; these depend on factors such as the neural connectivity and the firing rate function. It is also possible to consider how the set of attractors might change as the the various factors which influence the network shift. One example of this was the analysis performed by Brunel and Hakim (1999), who investigated how networks with different levels of inhibitory feedback have different attractor dynamics. In reality, the dynamics of the network are constantly modulated by external forces such as further sensory information, top-down attention (Larocque et al., 2014; Shamma et al., 2011), and the state of the physical body (the so-called 'Embodied Mind Thesis' (Varela et al., 1991)).

Dynamic Field Theory attempts to understand the interplay between these forces through bifurcation theory; the various influencing factors are understood as parameters which modify the location/stability of various attractors within the system (Sandamirskaya et al., 2013; Schneegans and Schöner, 2008). This area of study is only starting to be investigated in the literature, and is highly related to the concept of 'transient attractors' introduced in Chapter 2.

1.5 Models of Short-Term Memory

In this section we consider the challenge of short-term memory (STM) in networks of neurons. We start by defining various term and properties of STM, and then discuss

several models which have been proposed. This section provides a background to my work on STM presented in Chapters 2 and 3.

1.5.1 Nature of Short-Term Memory

In this dissertation we use the phrase Short-Term Memory (STM) to generally describe the ability to preserve information for short periods of time, typically from 100 ms to a few minutes. This is an essential property behind a wide variety of cortical computations. For example, reading this very sentence involves keeping track of various words and combining them into sensible meaning; hopefully the meaning itself perseveres, but the knowledge of individual words is quickly lost. Working Memory (WM) describes a specific type of STM that deals with the active focus upon and modulation of information stored in working memory while performing a task, such as navigating a maze (Dudai, 2002). However, not all papers make such a clear distinction, often using the two terms interchangeably (Mongillo et al., 2008; Szatmáry and Izhikevich, 2010).

STM has two limiting features. These are:

1. **Limited Duration.** For example, in tests of remembering separate items short-term memory lasts 15-30 seconds (Atkinson and Shiffrin, 1971). This can be consciously extended by repetitively attending to each object ("P Sherman, 42 Wallaby Way, Sydney, P Sherman, 42 Wallaby Way, Sydney, ...").

2. **Limited Capacity.** Strictly limited independent of task according to classic

"Seven, Plus or Minus Two" theory (Miller, 1956). Although the concept of some exact number has been challenged (Doumont, 2002), possibly even originally intended as little more than a joke (Miller, 1989), but there is a general agreement that capacity is highly limited. The memory capacity can be expanded by grouping objects together.

1.5.2 Experimental Evidence of Short-Term Memory

The exact location in the cortex in which short-term memory is stored is an area of active debate. Properties of short-term memory are often tested through the delayed recognition task (Figure 1.9), or similar tests. In this task, the subject must store some piece(s) of information for a period of time if they are to receive a later reward. Early experiments found that a brain region known as the prefrontal cortex (PFC) exhibited continued activity for the duration of such tasks (Fuster and Alexander, 1971; Kubota and Niki, 1971). This activity was later shown to be encoding task-relevant information (Courtney et al., 1997; Zarahn et al., 1997). This apparently indicated that the information was stored in continued activity, leading to the idea that STM is maintained in certain specialized brain regions by persistent neural activity.

A series of more recent experimental results, however, have drawn doubt upon such a mechanism underlying all short-term memory (Stokes, 2015). Firing rates are far lower, and far more variable, than expected from a standard persistent attractor

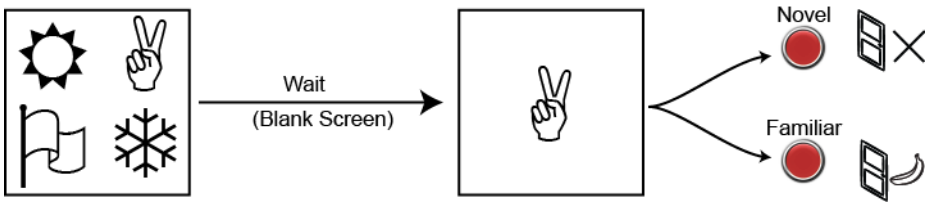

Figure 1.9: **Delayed recognition task for STM.** A standard test of short-term memory. A set of objects are shown to subject, followed by a delay period. Later, one object is shown, and subject must decide if this object is novel or familiar.

(Shafi et al., 2007). Moreover, should the subject know in advance when a memory is required, firing rates reflect the memory only immediately before it is required (Barak et al., 2010; Watanabe and Funahashi, 2007). In addition, such a method of memory storage is energetically expensive (Attwell and Laughlin, 2001) and may interfere with the network performing other tasks (Curtis and Lee, 2010).

Information storage might also occur within the sensory cortices themselves (Pasternak and Greenlee, 2005; Postle, 2015; Weinberger, 2012). For example, Petrides (2000) investigated the properties of short-term memory in monkeys with lesions in either parts of the PFC, or parts of the primary sensory cortices. They found that lesions in the sensory cortices influenced how long memories could be maintained, but not how many memories could be simultaneously maintained, whereas lesions in the PFC had the opposite effect. Other experiments have suggested that persistent activity does not represent all information, but instead information pertinent to the immediate mental processes (the so-called 'Focus of Attention') (Lewis-Peacock et al., 2012; Olivers et al., 2011). This suggests an alternative model in which the PFC provides executive control over the memories stored in the sensory cortices

(Postle, 2015).

We next consider various models of short-term memory, classifying them according to whether information is stored in the pattern of activity or in the network connection strengths.

1.5.3 Persistent Activity Models of Short-Term Memory

The standard model of short-term memory is one using a fixed point attractors (Barak and Tsodyks, 2014), as demonstrated in earlier discussions on the Hopfield Network (Section 1.4.2.1.1). Such a persistent attractor mechanism both predicts and depends on the persistent neural activity, as described above.

Self-sustenance is typically achieved in neural networks through different combinations of recurrent excitation (Goldman, 2009), inhibition (McDougal, 2011), or both (Aksay et al., 2007; Lim and Goldman, 2013; Machens et al., 2005). Variations on these models have discussed issues such as multi-item memory (Edin et al., 2009; Wei et al., 2012) or in cross-regional networks (Dubreuil and Brunel, 2016; Verduzco-Flores et al., 2009). Other models have proposed different forms of self-sustaining activity such as cycle attractors (Lisman and Idiart, 1995) and purely feed-forward circuits (Goldman, 2009). Finally, it has been suggested that short-term memory might be stored in 'dynamic coding' within the network (Stokes et al., 2013), which is essentially a chaotic attractor or a periodic attractor with very long orbits; no model has shown how information may be stored or reliably recalled in this way.

One interesting subset of papers has investigated how persistent neural activity due to a fixed point attractor might be influenced by transient synaptic modifications. In particular, they have shown how such networks might extend the duration (Barak and Tsodyks, 2007) and stability (Itskov et al., 2011) of the information, while also increasing number of memories which may be simultaneously stored (Rolls et al., 2013). These models share attributes with 'activity silent models' (covered below in Section 1.5.4) in their use of transient plasticity; I have chosen to classify them as persistent activity models as they still depend upon some continued neural activity.

Although the above models vary greatly in the underlying mechanisms, all store information within self-sustaining patterns of neural activity. Such a process is often credited with the observed delay-period activity during working memory tasks in the PFC. Nevertheless, for the reasons mentioned in Section 1.5.2, these processes are unlikely to fully explain the short-term memory faculties of the mammalian cortex.

1.5.4 Activity Silent models of Short-Term Memory

An alternative mechanism that might allow for the storage of information is temporary modifications to the effective connectivity to the network. This in turn changes the attractor structure of the network. In these models, ongoing information storage is not dependent on continued activity, but it may modify the characteristics of any ongoing firing patterns. Here we examine models which have been are capable of

transiently storing information solely within synaptic modifications.

1.5.4.1 STM from Transient Associative Modifications

Information about recent patterns of neural activation might be stored in short-term associative modifications to the networks connectivity. Such modifications temporarily strengthen the connections between neurons which are co-active, in turn meaning these neurons are more likely to be co-active in the future.

This method was first proposed by Buhmann and Schulten (1986). The neurons in this model were governed by Leaky Integrate-and-Fire dynamics (Section 1.2.1). Recurrent plastic connections exist between all neurons, governed by a transient associative rule. This causes positive connections to form between two neurons which are active at the same time, and negative connections to form between two neurons which are active at different times. In the absence of any activity, the weights return to their initial values.

When exposed to a pattern for a period of time, this network forms strong positive connections between pairs of neurons which both lie within the pattern and it forms negative connections between neurons which lie in the pattern and those that don't. It was demonstrated how this may lead to the completion of the pattern should it be shown later with some parts omitted. More recent work has extended this idea of short-term memory via transient associative plasticity. For example, Sandberg

et al. (2003) demonstrated how this mechanism could store regions of activity along a line attractor, and recall them in the presence of background noise. This idea has also been extended to periodic attractors, in which information is stored in a repeating pattern of neuronal activation (Szatmáry and Izhikevich, 2010).

1.5.4.2 STM via Synaptic Facilitation

Mongillo et al. (2008) demonstrated that short-term memory might also be stored using non-associative mechanisms such as facilitation (Section 1.3.2). This is only possible in a network with some underlying asymmetry in the connections between neurons; if the connectivity were perfectly homogeneous, facilitation of any individual neuron will influence all other neurons equally. The model presented by Mongillo et al. (2008) contained several non-overlapping groups of excitatory neurons embedded in a larger population of excitatory neurons. Connections between neurons in the same group were, on average, several times stronger than other connections (Figure 1.10A). The network also included a population of inhibitory neurons to mediate competition within the excitatory population.

A memory is stored in this network by stimulating one of the interconnected groups, causing the neurons within the group to facilitate. Although this facilitation is applied to all connected neurons, the asymmetries in connectivity means that this facilitation does not affect all neurons equally. Specifically, when a neuron is facilitated, it disproportionately excites neurons in the same group. The stored memory will

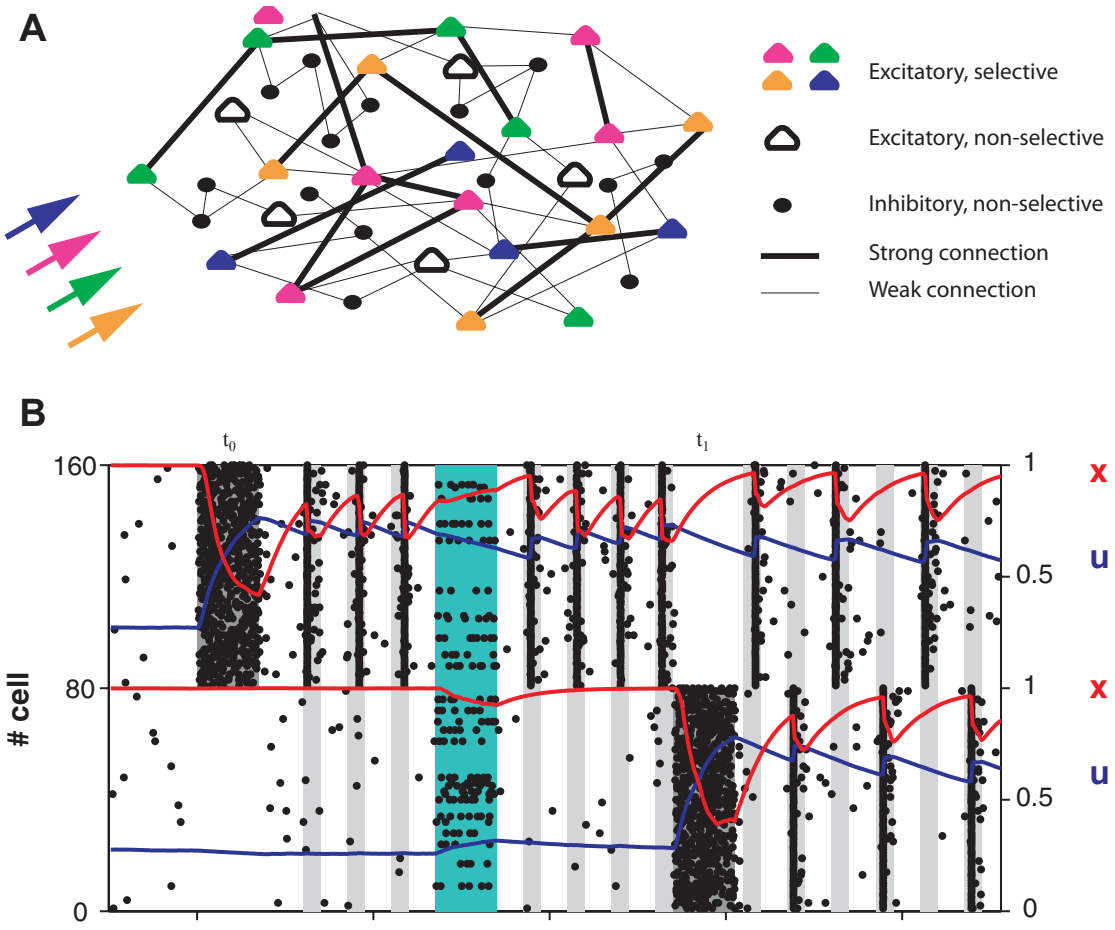

Figure 1.10: **STM via facilitation by Mongillo et al. (2008). A:** Network arrangement, with excitatory neurons divided into groups and stronger connections between neurons in the same group. **B:** Network behavior, showing spikes from a subset of neurons from two groups (1-80 and 81-160). Grey bars show periods of extra background noise, red line (x) records average depression in each group, blue line (u) average facilitation. Activity at t0 and t1 show external imprinting of memories. When one memory has been imprinted ($t_0 < t < t_1$), background noise causes retrieval of that memory. After a second memory is added later ($t > t_1$), noise prompts retrieval of one memory at a time.

then spontaneously reactivate in the presence of background noise (Figure 1.10B) due to these stronger connections. They also investigated how such a network might store two memories. This work demonstrates that facilitation can indeed store information in networks in an 'activity silent' manner; the network proposed, however, requires pre-wired groups to store each potential memory.

1.6 The Auditory System

Chapter 4 of this dissertation considers how the auditory system processes incoming sounds. In this section, we cover some background about how information is encoded in sounds, the steps involved in extracting that information, and how this information is represented within the auditory cortex.

1.6.1 Sound generation

In order to understand how sound can be analyzed, it is instructive to consider its generation. Many sounds are produced by acoustic oscillators, such as a vibrating vocal chord or a plucked guitar string. These produce sound waves which are more or less periodic. These periodic waves can be decomposed into a harmonic stack, i.e. a sum of sinusoids with frequencies that are integer multiples of some fundamental frequency. Not all harmonics are represented equally; sound generation may lead to emphasis on certain harmonics (e.g. reed instruments typically contain only odd-numbered harmonics), or emphasis at certain frequencies (known as formants; "ee" vowel contains energy peaks at 280 Hz, 2300 Hz and 2900 Hz). This is illustrated

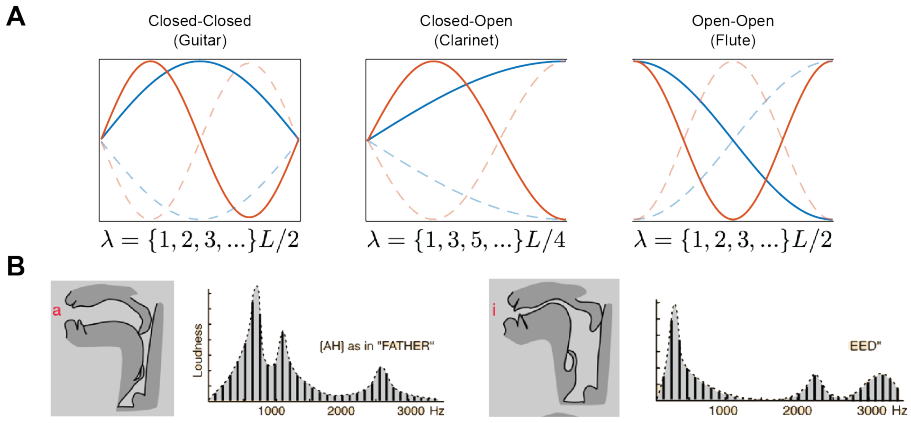

Figure 1.11: **Harmonics and formants. A:** The nature of the acoustic oscillator influences which harmonics are be generated (black bars), a key component in the timbre of a sound. **B:** The shape of the vocal tract dictates which ranges of frequency are accented or muted. Peaks in energy are known as formants, and the nature of a vowel is highly dependent upon their locations. Unvoiced speech maintains the general energy distribution but loses harmonics (i.e. grey shading under curves without the black bars).

in figure 1.11.

Not all sounds, however, are generated from continuous vibrations in a medium of some kind. Examples of such non-harmonic sounds include several consonants and unvoiced speech (whispering). In my dissertation, however, I concentrate on the processing of harmonic stimuli.

1.6.2 Detection and Representation of Sounds

Next we consider how the auditory system detects and responds to auditory stimuli. This process starts at the cochlea which detects the amount of energy present at various frequencies (20 Hz to 20 kHz). The phase of the components is also recorded for low-frequency waves (under 3 kHz); this information is used in sound localiza-

tion using inter-aural time difference. This information is then transferred to the primary auditory cortex via the auditory nerve.

Individual neurons in the primary auditory cortex respond to energy at a certain frequency, and these neurons are arranged tonotopically according to this frequency. These neurons are also known to be sensitive to various other, higher order features. These include the onset and offset properties of the sound (Qin et al., 2007), the distance between harmonics (Bendor and Wang, 2005) (known as 'spectral pitch'), gradual shifts in spectral composition (Theunissen and Elie, 2014), apparent location of the source (Oswald et al., 1999; Rose et al., 1966), and several other factors. How these various elements are combined to allow for a coherent auditory percept is not well understood.

1.6.3 Auditory Tasks

1.6.3.1 Identification and Localization

Perhaps the most obvious reason to process incoming sound waves is to determine information about the process which generated the sound. Such information may be deliberately encoded in the sound (e.g. mating calls including the pop hit "Gangnam Style") or incidental (a predator's footsteps, a falling branch). In order to consider the various ways in which information is encoded in sound, it is enlightening to consider the language that has developed to describe sounds (Table 1.1). We note that the term 'timbre' is generally used to refer to the combined effect of several

effects such as harmonics, coloration, and temporal envelope.

1.6.3.2 Auditory Streaming

The auditory landscape typically includes sounds from multiple different sources; for example, the sound at a party may be a mix of several individual voices along with music, traffic, and several other background noises. The ease with which both humans and animals may separate this scene into its constituent parts belies the difficulties within such a calculation. We refer to this process as auditory streaming, that is the separation of an input signal into its constituent streams. Streaming is highly linked to, but slightly different from, the so-called 'cocktail party effect' (Cherry, 1953), which is concerned with the separation and amplification of one particular stream from an input composed of several sources.

The precise definition of an auditory stream (or of its constituent auditory objects) remains much debated (Bizley and Cohen, 2013). However, an important point of agreement is that they are perceptual; streaming describes how the auditory system interprets the input signal. We might hope that each auditory object represents an auditory event, or each auditory stream represents an auditory source, but this is due to the functioning of the cortex as opposed to being inherent in the definition.

The factors which impact the division of auditory input into streams include:

- **Frequency differences** (Miller and Heise, 1950). Alternating between two

Name	Cause
Volume	Amplitude
Tremolo	Oscillations in amplitude
Pitch	Fundamental frequency
Vibrato	Oscillations in fundamental frequency
'Clean' sound	Fraction of energy in harmonics
'In tune'	Simple ratios between harmonics present (e.g. 3:2 = 'Perfect Fifth')
Rhythym	Long scale periodicity(>100ms)
Coloration/Formant (e.g. different vowels)	Spectral envelope
Coloration/Formant glide (e.g. dipthong)	Changes in spectral envelope
Attack/Decay/Sustain/Release	Temporal envelope
Where (left/right)	Interaural time difference
Where (left/right)	Interaural intensity difference
Where (all)	Filter resonance from body and outer ear
Where (all)	Dynamic cues (deliberate head/ear movements)

Table 1.1: Examples of perceived features, and their underlying causes (Erickson, 1975; Haykin and Chen, 2005)

tones (a trill) is perceived as two separate streams if the two notes are sufficiently different in frequency

- **Temporal similarity** (Schnupp et al., 2011). Minimizing the time difference between tones increases the probability they will be perceived as a single stream

- **Timbre** (Wessel, 1979). Notes with a similar timbre are associated into a single stream

- **Spectral Continuity** (Darwin and Bethell-Fox, 1977) Alternating vowels may be perceived as one or two streams depending on the exclusion or inclusion of intervening glides

- **Spatial Location** (Deutsch, 1979). One melody played from multiple places will not be perceived as a single stream

- **Pitch** (Brokx and Nooteboom, 1982). When differences in pitch are removed from multiple voices (both voices are modified to be monotone at the same pitch), it is increasingly difficult to separate voices

- **Attention** (Carlyon et al., 2001). Attending to competing tasks can prevent stream segregation

- **Fine spectral-temporal features** (Ding et al., 2014; Qin and Oxenham, 2003). Removing fine spectral-temporal structure does not affect comprehension of a lone speaker, but makes it far more difficult to attend to a speaker if another distraction speaker is present.

See Darwin (1997) for further discussion on the factors influencing streaming.

As streams are perceptual, the streaming experiments typically ask the subject to react depending on what was perceived. Nevertheless, stream segregation must occur relatively early in cortical processing. For example, subjects have great difficulty determining the relative timing of events when they are perceived to occur within different streams (Bregman and Campbell, 1971).

1.7 Models of Auditory Streaming

Chapter 4 of this dissertation presents a novel, neural network approach to explain the process of auditory streaming. Consequently, in this section I will provide a review of different mechanisms which might be involved in this process. I do so using the three key aspects of streaming as proposed by Bregman (1990): segmentation, segregation, and integration. I will then review some neural network models which have been suggested to explain auditory streaming in cortex.

1.7.1 Segmentation

Segmentation involves examining the instantaneous auditory information represented in the waveform, and deciding which parts should be combined together into distinct auditory objects (Figure 1.12). This will involve analyzing the sound presented along multiple different dimensions (Table 1.1). It is also known that these features are indeed represented by the neural code (Section 1.6.2), and that

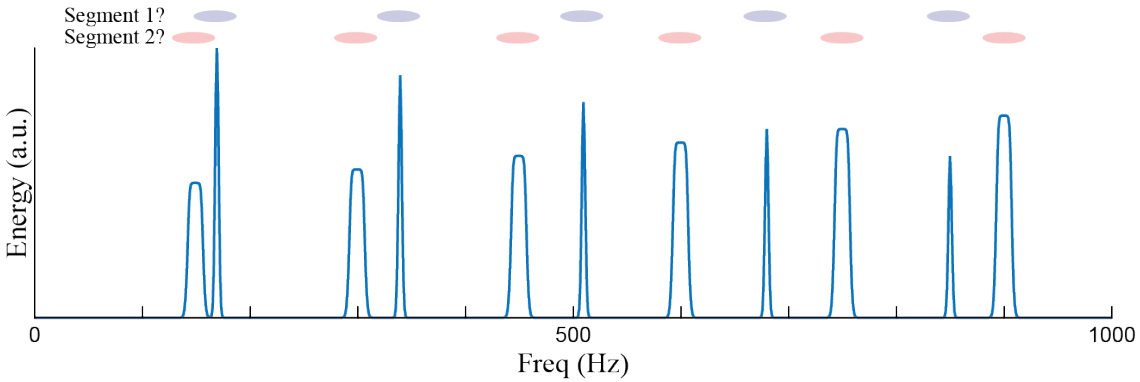

Figure 1.12: **Auditory segmentation.** Snapshot of frequencies over time may allow the sound to be separated into distinct segments using multiple features such as periodicity over frequencies (pitch), width of individual bands (warmth) and shape over multiple bands (color). Possible clustering into segments shown by colored circles above the plot.

different parts of the auditory stimulus are more likely to be perceived as a single object if they match in several respectes (Section 1.6.3.2). This makes intuitive sense; sounds generated from one object typically share multiple features, and so forming streams with multiple shared features allows the system to label features from the different sources. Such a decomposition of auditory stimuli according to clusters of shared features is popular in the CASA literature (Brown and Cooke, 1994; Elhilali and Shamma, 2008; Wang, 1996).

1.7.2 Segregation

Segregation describes the network's ability to simultaneously represent the multiple objects detected through segmentation. This becomes particularly challenging when we consider that each incoming sound is decomposed into multiple different

features; somehow, the cortex must store a label for each feature to know to which auditory object it has been allocated. This challenge is also known as the binding problem, i.e. how the cortex tracks which features are currently 'bound' into a single perceptual object (Roskies, 1999; Von der Malsburg, 1981).

One of the most widely discussed solutions to auditory segregation is correlation theory, which suggests that neurons which represent different features of each object will typically fire synchronously with one another. This synchronicity is thought to be related to the various cortical oscillations occurring within biological neural networks; it is hypothesized that subpopulations of neurons firing in phase during such oscillations are 'bound' together. Evidence for binding via synchronicity has been most thoroughly investigated in the visual cortex, in which synchronous populations of neurons have been found to be associated with the same object (Fries et al., 1997; Singer and Gray, 1995). Such synchronous populations of neurons have also been observed in several different parts of the sensory cortex (Engel and Singer, 2001), including the auditory cortex (Ehret, 1997), although the role that these might play in object segregation is highly contested (Roskies, 1999).

1.7.3 Integration

The third step in auditory streaming is integration; that is, joining together the instantaneous segments into a single, continuous, perceptual stream. Experimental evidence suggest that this process is the last to happen, after both segmentation and

segregation (Sussman, 2005). Several different mechanisms have been hypothesized to be involved in this auditory integration.

One possible mechanism for integration is by assuming a continuous neural representation of each stream. This is due to the fact that many features that come from one source of sound only evolve slowly. For example, even while the format structure and pitch of a voice may shift quickly during conversation, the timbre and location remain mostly constant. Psychoacoustic experiments have indeed confirmed that the continuity within the source will influence the number of streams perceived (Darwin and Bethell-Fox, 1977)(Figure 1.13A).

Algorithms have been proposed which form streams by maximizing continuity of features within each stream. For example, Elhilali and Shamma (2008) (Figure 1.13B) suggested an algorithm using the concept of a Kalman filter (Kalman, 1960). This algorithm is cyclical; incoming sounds are segmented according to the perceived nature of the streams, and the streams are then updated using these segments. One significant advantage of this process is that calculations may be done online, with new segments being instantly joined to those streams which they bore the greatest similarity.

An alternative approach to integration is to use a property known as temporal coherence. When a source generates a sound, it may output multiple components (e.g. harmonics). However, the various components of the sound are all generated by the

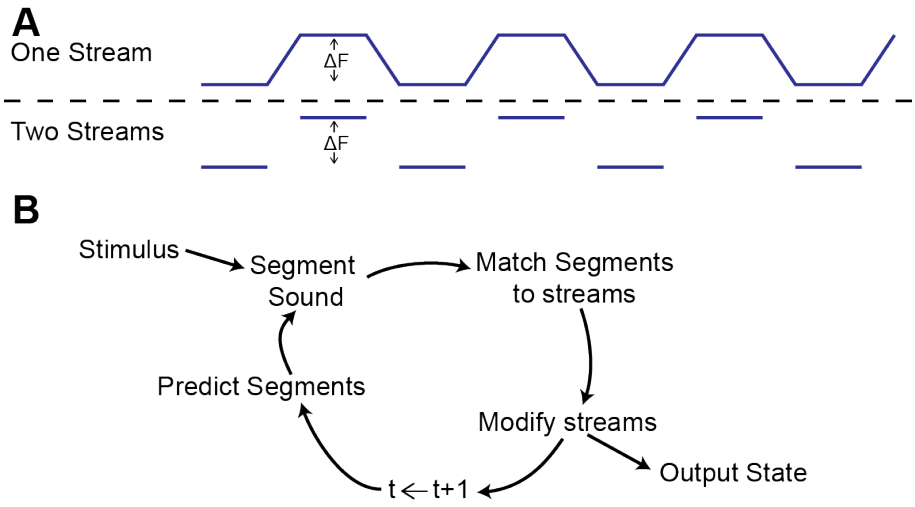

Figure 1.13: **Continuous features perceived as a stream. A.** Continuous deformations to a sound's features (*top*) are more easily bound into a single stream than 'jumps' of the same size (*bottom*). **B.** Algorithm presented in Elhilali and Shamma (2008) for joining segments into streams by maximizing continuity between successive segments. Based on Kalman filter.

same underlying cause, and so are likely to share some temporal envelope: that is, they start at the same time, undergo similar and simultaneous changes in strength, and finish together. This idea underlies integration via temporal coherence; features which have similar temporal envelopes most likely were generated by the same process. Each stream is decomposed into its component features and a temporal envelope. Moreover, these properties of sound have been shown to influence the perception of streams (Teki et al., 2013).

Several models have been described which derive integration via temporal coherence type. Variants have used Independent Subspace Analysis (Casey and Westner, 2000), non-negative matrix factorization (Smaragdis, 2004) and autoencoders (Krishnan et al., 2014). One issue with these modeling approaches is in identifying the

features; many models simply assume that the features are the inputs, so that all the features must be stationary across time. Unfortunately, features are typically nonstationary. For example, consider the changes to a harmonic's location as the underlying pitch changes. Smaragdis (2004) suggested a solution that would allow features to be defined over time; the proposed algorithm could then detect moving features, as long as they all move in exactly the same manner. An alternative approach was suggested by Krishnan et al. (2014), who constructed a series of correlation matrices over a variety of frequency and temporal filters. These correlation matrices are then decomposed using an autoencoder. In this way, objects which are sufficiently close in frequency and time are associated together. Such an algorithm reproduced some experimental findings concerning streaming, however it is unlikely that such a calculation could be performed by a neural network.

1.7.4 Neural Networks for Auditory Streaming

We finish this section by reviewing two neural networks which have been proposed to solve some of the above challenges associated with auditory streaming.

1.7.4.1 "A Neural Cocktail-Party Processor"

Shortly after the paper introducing the binding problem and correlation theory (Von der Malsburg, 1981), Von der Malsburg and Schneider (1986) proposed a neural network that could separate auditory stimuli into distinct streams. The network proposed was intended as a sketch; consequently, it does not discuss particular cel-

lular mechanisms or timescales. The basic units in the model are neural oscillators; when connected to a persistent input (all inputs being binary), each oscillator automatically alternates between periods of activity and inactivity.

In this model, segmentation is performed using transient associative plasticity; if two inputs are present at a similar time, the connection between them increases in strength, meaning they are more likely to be co-active in the future. Segregation then occurs in keeping with correlation theory due to these strengthened connections — paired units are observed to start oscillating in in time with one another. The network also contains inhibition-mediated competition, ensuring that multiple perceptual objects will oscillate out of phase with one another. In this model, the auditory sources emit a constant sound, so there is no need for integration to occur. This model demonstrates how a network of neural oscillators might achieve segregation using learned correlations.

1.7.4.2 Local Excitatory Global Inhibitory Oscillator Network (LEGION)

The concept of streaming occurring through a network of neural oscillators was extended by the Local Excitatory Global Inhibitory Oscillator Network. This model was first proposed as a solution to the binding problem in 1995 (Wang and Terman, 1995), and has since been investigated in a series of papers related to auditory streaming (Brown and Wang, 1997; Shao and Wang, 2009; Terman and Wang, 1995;

Wang, 1996; Wang and Chang, 2008; Wang and Brown, 1999; Wrigley and Brown, 2004). As with the Neural Cocktail-Party Processor, the basic units are neural oscillators (highly reminiscent of the van der Pol oscillator) which undergo alternating periods of activity and inactivity. These units each represent one particular frequency, possibly also with a certain time lag. Recurrent connections exist between 'close' units, that is those units which represent similar frequencies (and possibly also similar time lags) (Figure 1.14).

In this case, segmentation is performed due to the pre-established recurrent excitatory connections: the units which are connected are more likely to oscillate in time. The key difference from von der Malsburg's work above is that these associations are pre-encoded, rather than learned by temporal coherence. This allows for some memory in the network that informs what features should be 'bound'. Segregation then occurs due to the oscillating nature of the network, in keeping with correlation theory. Finally, integration with the inclusion of the lagged inputs: if two features are close in frequency space, but are not present at the same time, they may still be associated together using a time-lagged version of the first feature. This model demonstrates how a recurrent neural network might achieve the segmentation, segregation, and integration hypothesized to underlie auditory streaming.

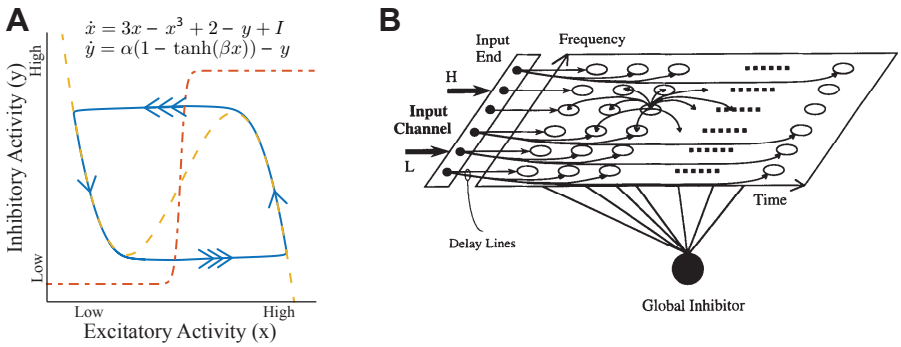

Figure 1.14: **The Local Excitatory Global Inhibitory Oscillator Network. A:** Oscillator alternates between periods of high and low excitatory periods, with fast transitions between the states. **B:** Network structure, with local excitatory connections across frequency and time

Chapter 2: Cortical Computations via Transient Attractors

2.1 Overview

The ability of sensory networks to transiently store information on the scale of seconds can confer many advantages in processing time-varying stimuli. How a network could store information on such intermediate time scales, between typical neurophysiological time scales and those of long-term memory, is typically attributed to persistent neural activity. In this chapter, I will describe an alternative, a Transient Attractor network, supported by synaptic plasticity between individual connectivity that decays on the same second-long time scale. I hypothesize that such functionality could be usefully embedded within sensory cortex, and confer not only aspects of short-term memory, but also automatically de-noising inputs, and separating them into distinct perceptual objects. From the perspective of short-term memory, this network can learn any pattern presented to it, store several simultaneously, and robustly recall them on demand using targeted probes in a manner reminiscent of Hopfield networks. The stored information can be refreshed to extend storage time, is not sensitive to noise in the system, and can be turned on or off by simple neuromodulation. Moreover, I will show how the same mechanisms may assist in the processing of sensory information on the seconds-long time scales. The diverse

capabilities of transient attractors, as well as their resemblance to many features observed in sensory cortex, suggest the possibility that their actions might underlie neural processing in many sensory areas.

2.2 Introduction

The real world "causes" of sensory inputs usually persist for much longer than the time scales of neural processing in sensory areas. As a result, there is great use for neural and circuit mechanisms within sensory cortex that can hold information over time scales much longer than neural time scales, on the order of seconds. Storage of information on this time scale is commonly addressed in the context of "short-term memory" (Maex and Steuber, 2009), but there are also other uses for seconds-long storage of information. For example, such aggregation of information over time can be used to segregate auditory stimuli into perceptual auditory objects (Krishnan et al., 2014). Similarly, features of visual objects can be assembled over time using such associations despite temporary occlusions and visual noise (Becker, 1992).

The most common models of short-term memory rely on the concept of a "persistent attractor" (Barak and Tsodyks, 2014; Seung, 1996). A fixed set of recurrent connections can support a number of different patterns of activity ("attractors"), whereby the activity patterns are self-reinforcing due to excitatory connectivity. Persistent activity is typically maintained by a combination of excitatory and inhibitory activity (Aksay et al., 2007; Machens et al., 2005), and persistent states can even exist

in random networks with particular properties (Ganguli et al., 2008). The unifying feature of persistent attractors networks is that information is stored in neural activity itself, thus keeping it readily accessible.

Although there is some evidence of information storage in persistent attractors in specialized brain regions such as the pre-frontal cortex (D'Esposito and Postle, 2015), it is unlikely to explain short-term memory in other areas such as the sensory cortices (Pasternak and Greenlee, 2005; Petrides, 2000; Postle, 2015), and perhaps not even explain all information storage within the specialized regions (Barak and Tsodyks, 2014; Sreenivasan et al., 2014; Stokes, 2015). An alternative location for the storage of information about recent inputs is in the local connectivity within the network itself. Indeed, such memory storage is implicit in models of long-term memory (Hopfield, 1982), where memories are encoded in the excitatory connectivity which is established using a simple form of associative plasticity. Such a scheme could also be used for short-term memory if such changes in synaptic connectivity were temporary, which would allow for the short-term preservation of information within the network without affecting the network's long-term structure (Buhmann and Schulten, 1986; Mongillo et al., 2008; Sandberg et al., 2003; Szatmáry and Izhikevich, 2010). This allows the network to return to a certain state of activity (an attractor) in the presence of appropriate inputs (Schneegans and Schöner, 2008). I label such attractors "transient" as they only exist during appropriate input and due to relevant changes to network connectivity (which are themselves temporary).

In this chapter, I will propose transient attractors as a unifying mechanism within cortical networks that can support multiple types of computation that require combining information across time scales much longer than those of the underlying neurons. I will first demonstrate how a transient attractor functions in the context of a classic short-term memory task. Several memories can be stored in the network structure, allowing for their recall in the presence of suitable inputs. These memories then fade over several seconds. The same network can be used to extract information from time varying stimuli, specifically in the tasks of stream segregation and signal de-noising. I finish by considering some issues that impact the various uses of transient attractors, including transient attractor maintenance, the effect of top-down attention and the overall robustness of the network.

2.3 Results

I considered a simple form of a transient attractor network, composed of a single layer of excitatory and inhibitory neurons (Figure 2.1A), and in which neural activity is represented with firing rates governed by leaky integrate-and-fire dynamics. Excitatory neurons receive feedforward inputs representing the stimulus, as well as recurrent connections. For this simple network, only a single inhibitory unit is required. This neuron receives inputs from, and projects back to, the excitatory neurons and itself. This inhibitory circuit mediates competition between ensembles of excitatory neurons to represent the input. This minimal circuit can produce the behavior described below, but can both be expanded into a larger network, as well

as be produced by many other variants (see Discussion).

Short-term memory in this model is supported via plastic recurrent excitatory connections. The total synaptic strength of connections between each pair of excitatory neurons is the product of three terms: a baseline synaptic weight S_{ij} between neurons i and j, an associative (Hebbian) gain H_{ij} that can be increased by coincident pre- and postsynaptic activity, and a synaptic depression x_i, governed by the equations

$$W_{ij}(t) = S_{ij}H_{ij}(t)x_i(t) \tag{2.1}$$

$$\frac{dH_{ij}(t)}{dt} = \frac{1}{\tau_{H_+}}\left[H_{max} - H_{ij}(t)\right] y_i(t)y_j(t) - \frac{1}{\tau_{H_-}}\left[H_{ij}(t) - H_{min}\right] \tag{2.2}$$

$$\frac{dx_i(t)}{dt} = \frac{1}{\tau_{x_+}}\left[1 - x_i(t)\right] - \frac{1}{\tau_{x_-}}\left[x_i(t)y_i(t)\right]. \tag{2.3}$$

I assume that all such recurrent connections have a uniform baseline strength ($S_{ij} = S_0$), and connections to and from the inhibitory neuron will also be assumed to be uniform. As I will describe below, this gives the network the maximum potential for memory storage, but alternatives will be considered below.

Excitation is regulated (and stable) due to two mechanisms: feedback inhibition and synaptic depression. The inhibitory circuit suppresses all neurons by an amount proportional to the total excitatory activity, resulting in competition between the excitatory neurons. Synaptic depression controls recurrent excitation within a neuron group, limiting the dominance that persistent attractors can have so that activity is more directly influenced by current stimuli (as filtered by recent history). Further

details about the model can be found in the Methods.

In the presence of a constant external input, firing rates in the network will settle into some pattern (an attractor) that depends on both the external input and the state of the network. Note that this definition of an attractor is broader than that used in much of the persistent attractor literature, which only considers attractors that remain active when external input is removed. The dominant attractor at any moment will depend on both the stimulus present and the effective synaptic strengths, which in turn depend on recent history of network activity through the associative gain term (H_{ij}).

2.3.1 Short-Term Memory via Transient Attractors

I start by illustrating how the transient attractor network works within a minimal network with just four excitatory neurons (Figure 2.1A). I selected two patterns to store: the first with neurons #1 and #3 coactive, and the second with neurons #2 and #4 coactive (Figure 2.1B). Before the memory is stored, I presented "probe" stimuli, each driving a single neuron (Figure 2.1C, *left*) in order to verify there is no preexisting network attractors. Indeed, such probe stimuli only evoke activity in the neurons that were externally stimulated (Figure 2.1D, *left*). To imprint the memory, the two patterns are displayed alternately at 4 Hz for 1 sec (Figure 2.1C, *center*). Following this, both probe stimuli are displayed again (Figure 2.1C, *right*)

to determine if the memories are recalled in the network activity. Indeed, while only the stimulated neurons fire in response to the probe stimuli at the beginning (Figure 2.1D, *left*), the patterns emerge after training (*right*).

During the training period the memory is imprinted in the increased recurrent weights between co-active neurons over repeated presentations (Figure 2.1E). These strengthened connections then lead to increases in membrane voltages in response to even a part of a pattern in which it has been involved (Figure 2.1F). This in turn causes an increase in inhibitory firing rates proportional to the additional excitatory activity (Figure 2.1G), and an increase in suppression of the non-paired neurons. These inhibitory effects are small relative to excitation due to the size of the pattern being stored in this small network, and in general will have larger effects that create competition between simultaneously active neurons.

I next extended this simple example to a much larger network, capable of learning multiple, overlapping patterns. This network has 100 excitatory neurons, arranged in a 10 by 10 grid. Note that the grid arrangement is only to make visualizing the patterns of activity easier, and it does not represent any biases in connectivity; again, all excitatory connections are all-to-all, and of equal strength. I trained this network with three patterns, two digits (to be easily recognizable) and a third composed of randomly selected neurons. This set of patterns illustrates how any pattern can be stored in the network, but also note that the two digits chosen have a large number of shared elements. Random subsets of each pattern are selected

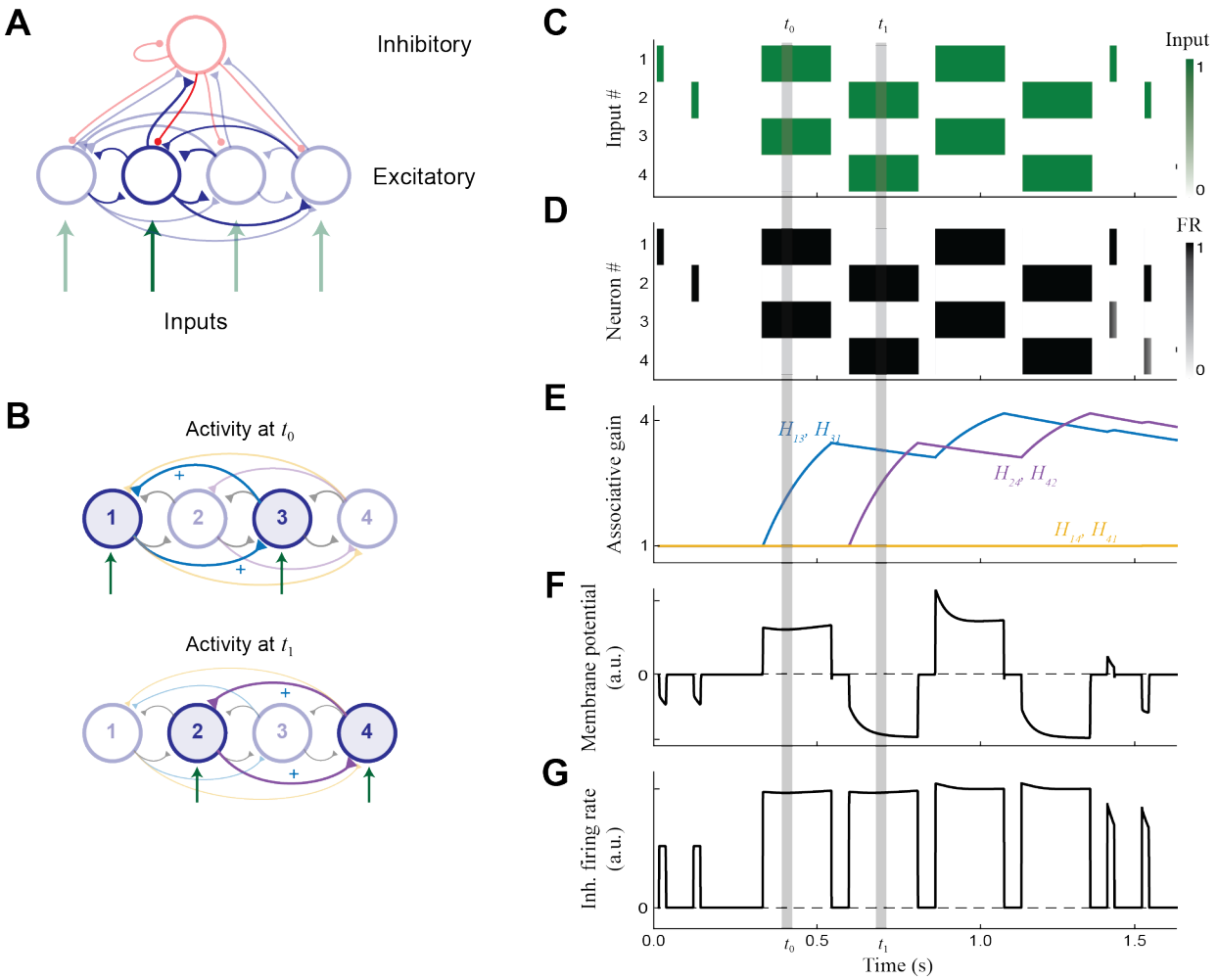

Figure 2.1: **Transient attractors in single layer network via associative weight modifications.** **A.** Diagram of a minimal transient attractor network composed of four excitatory neurons and one inhibitory neuron. Recurrent connections between excitatory neurons are plastic, and all connections to/from the inhibitory neuron are equal. **B.** *Top*: When presented with inputs to neurons 1 and 3, recurrent connections between simultaneously active neurons are strengthened through transient associative plasticity. *Bottom*: A second stimulus (inputs 2 and 4) is stored through the strengthening of recurrent connections. **(C-G)** Two stimulus patterns are stored and recalled by the minimal transient attractor network in (A). **C.** The inputs to the network, composed of: a probe stimulus activating single neurons (*left*), followed by two repetitions of two different two-neuron patterns (shown in B), followed by the same probe stimulus (*right*). **D**. Activity of excitatory neurons in response to stimulus. **E.** The synaptic gains H_{ij} strengthen as a result of the coincident neural activity, shown for a representative sample of connections. **F.** Membrane potential of neuron #3. Initially, both probes cause some inhibition, while after training the in-pattern probe causes elevated potential (firing), while other probe causes increased inhibition. **G.** Inhibitory neuron's firing rate, which increases when any pattern is presented. Resulting competition between patterns suppress one pattern in favor of the other when final probe stimuli are presented (*right*).

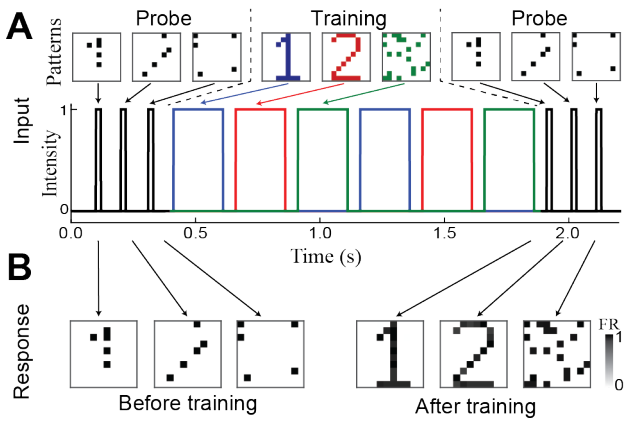

Figure 2.2: **Transient attractors can simultaneously store several patterns.** A network of 100 excitatory neurons stores and recalls three different patterns. **A.** Input to the network: *Center*: three patterns are each presented twice during training. *Left/Right*: Probe stimuli before and after training, each probe random subset (25%) of relevant pattern. **B.** Neural activity (firing rate represented in grayscale) only recapitulates the probe inputs before (left), but separately recalls each pattern (right) after training.

as probe stimuli, and the network is tested to have no pre-existing attractors, and trained as described above (Figure 2.2A). The successful storage of the memories in the network can be verified by comparing the levels of activity of the excitatory neurons to the initial and final probes (Figure 2.2B). This shows that an attractor has been created for each pattern. Furthermore, due to the inhibition-mediated competition, activity does not leak between overlapping attractors, and the stored information is recalled in the presence of a relevant probe. This demonstrates that this network is capable of performing short-term memory tasks involving multiple (potentially overlapping) memories held simultaneously. As with Hopfield networks, the memory capacity of this network (i.e., the number of patterns that can be stored simultaneously in memory) increases with the number of neurons (Amit et al., 1985), but in practice such a capacity cannot realistically be used due to the limitation of

the transient time scale over which the trained patterns of connectivity maintain themselves.

Stored short-term memories in this network have an additional attractive property in contrast to persistent-activity-based attractors: namely that they are stable while being stored. Such stability can be demonstrated in an example network where there is a clear topography between different activity states of the network. Thus, I next consider a ring attractor (Ben-Yishai et al., 1995). A ring attractor is composed of a circle of neurons, with each neuron preferentially connected to its neighbors (Figure 2.3A). In principle, ring attractors based on persistent activity can store a continuous variable because activity at any point on the ring can be stable. However, it has been shown that any noise in recurrent connections will cause a severe reduction in the number of stable equilibriums: typically down to a handful (Itskov et al., 2011). In practice, this means that the system will always drift to one of the relatively few global attractors (Figure 2.3B).

Transient attractors avoid this drift by having the network inactive in-between training and read-out (Figure 2.3C), meaning that the memory cannot drift. This observation complements earlier work (Itskov et al., 2011) showing plastic synapses will reduce the rate of drift in the case of persistent activity (Figure 2.3D). Furthermore, analogous to the more general network considered above (Figure 2.2), this network is capable of storing multiple locations simultaneously (Figure 2.3E), each re-activated by their own probe. This demonstrates how storing information in modified synap-

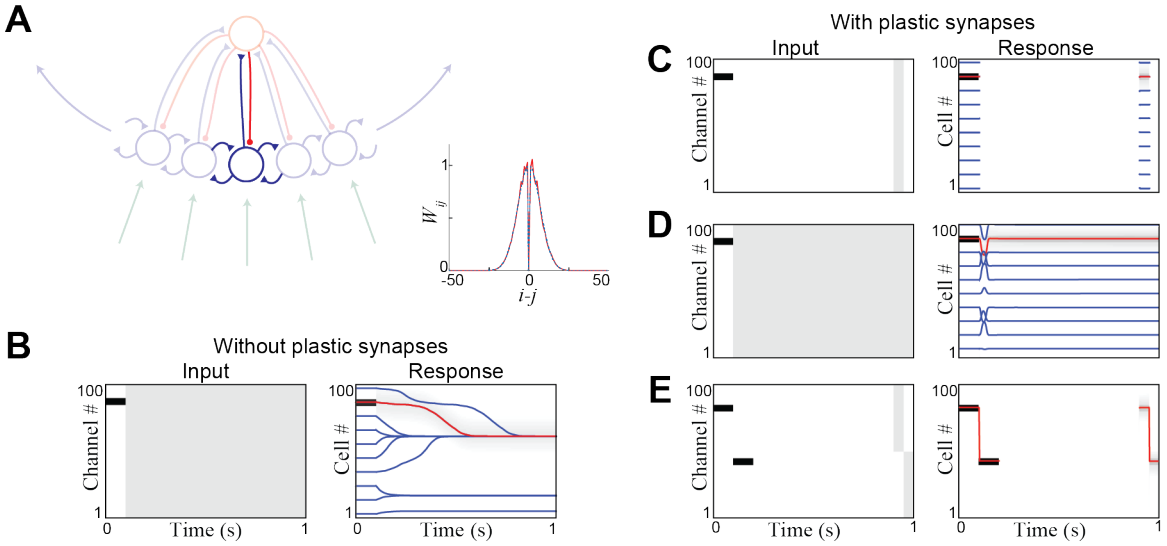

Figure 2.3: **Short-term memory in a ring attractor. A.** The network structure of a ring attractor, showing a section of the full ring. Each excitatory neuron (blue) connects to its nearest neighbors (with a Gaussian weight profile, see *inset*) and to a single inhibitory unit (red). **B.** After initializing a standard ring attractor to a single input (*left*) and then giving weak uniform input to keep the network active (gray), the position of the peak activity (*right*) drifts, often converging to one of several attractors. The trajectories of peak activity for several different initial inputs are shown. **C.** A ring attractor with transient synaptic plasticity allows for information storage in transient attractors at any single location. In this case, once information is first stored (ten initial inputs shown, *left*), there is no activity in the network (*right*). However, when probed with nonspecific uniform input, the network recalls the initialization with no drift. **D.** Transient attractors stabilize activity in case of persistent activity (ten initializations as before). The transient departure from the initialization is due to synaptic depression, but stabilized after the synapses recover. **E.** Transient attractors also allow for simultaneous storage of multiple locations. Here, two positions are stored, and recalled with more specific but still broad activation.

tic connections, as opposed to persistent activity, prevents slow distortion of the information by small errors within the network (in this case, attractor drift).

2.3.2 Maintenance of Information over Time

By design, information stored in transient attractors degrades at the time scale of the underlying transient synaptic plasticity. While this would appear to limit the amount of time a memory can be stored by the transient attractor, such a network can extend to storage over longer periods of time through reactivation of the attractor (Mongillo et al., 2008). Such reactivation will strengthen all relevant connections, and thereby allow information to be stored for durations well past the time scales of the decay of the transient synaptic plasticity.

To demonstrate how the transient attractor is capable of this, I first stored two overlapping patterns (Figure 2.4A, *left*). Without any further activity, the information stored will become inaccessible over several seconds due to the timescale of decay of the induced synaptic plasticity. However, here the stored information is refreshed by regular reactivation of the attractors via pulsing background activity (Figure 2.4A, *center*). Background stimulation causing the refresh need not be specific to any stored pattern; in this example, background stimulation is uniform across all channels, but as a result momentarily activates individual attractors within the network. Furthermore, the pulsing nature allows for sequential activation of multiple attractors due to the synaptic depression of synapses which were most

recently activated. The pulsing uniform activity is not the only conceivable method of refreshing memories; for example, specific memories might be targeted using an appropriate probe. As a result of this attractor reactivation, it can be seen that the duration of the memories has been extended (Figure 2.4A, *right* and Figure 2.4B). This demonstrates how transient attractors could store information over variable time scales.

2.4 Associating Distinct Patterns of Input via Temporal Coherence

For the above examples of memory, stimuli were presented separately in time in order to focus on the storage and retrieval of patterns. However, real world stimuli will often not be so conveniently separated in time, with different components that can only be distinguished by detecting shared temporal features. Such a theory of temporal coherence has been suggested as a solution for the cocktail party problem, that is the ability to associate the features comprising different sounds and focus on those components while suppressing others (Bizley and Cohen, 2013; Shamma et al., 2013). Temporal coherence has likewise been used for visual object separation (Becker, 1992).

The network described above can perform a simple example of such segregation based on temporal coherence. I generated a training stimuli composed of two random, non-overlapping patterns of activation, modulated by uncorrelated temporal envelopes (Figure 2.5A). As with earlier examples, probes are displayed before and

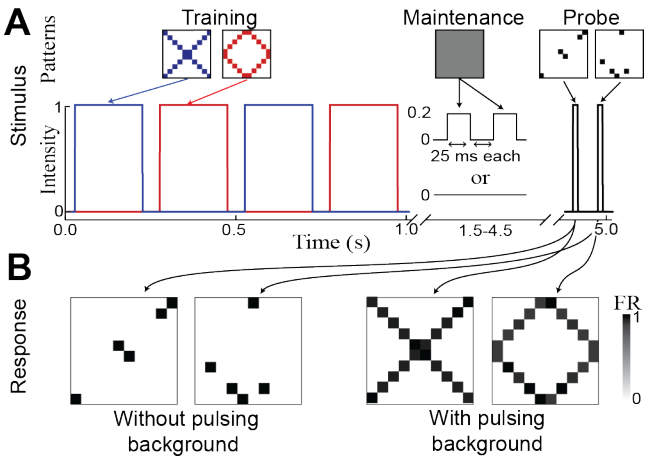

Figure 2.4: **Maintenance of transient attractor by uniform input. A.** The transient attractor network from Fig. 2 is initially trained with two patterns, and probed to test for memory storage almost 4 seconds later, well longer than the decay time of the transient plasticity. The intermediate period is either silent or consists of pulsing low intensity uniform network inputs (inset labeled Maintenance). **B.** The excitatory response to the probe stimuli either without (*left*) or with (*right*) the maintenance activity reveals how continued activity allows the network to maintain transient attractors and extends duration of information.

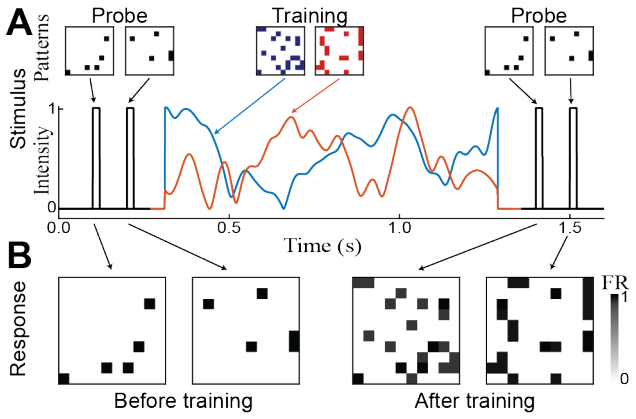

Figure 2.5: **Network can distinguish between patterns using temporal coherence. A.** The same 100 neuron transient attractor network from Figure 2 is trained on two patterns that are presented simultaneously, although with distinct temporally varying amplitudes (red, blue), with probe stimuli before and after, as with previous figures. **B.** The excitatory activation to the probe stimuli before (*left*) and after (*right*) reveal the ability to recall the two separate patterns, despite not being presented separately.

after exposure to patterns to demonstrate the creation of transient attractors. While both patterns were present at some amplitude throughout the training period, the network responses to the probes (Figure 2.5B) following training reveal that the network has learned both patterns. This happens due to the inhibitory feedback which prevents both patterns from being represented simultaneously. As patterns in the network aren't represented simultaneously (even if both are present in the stimulus), they are essentially temporally segregated within the network allowing associations to be learned. I conclude that the transient attractor network is capable of performing some form of on-line temporal coherence analysis, and completes a simple streaming task.

2.4.1 Separating Signal from Noise

Just as networks with persistent activity may act as neural integrators (Shen, 1989), the transient attractor network may also act as an integrator, allowing it to filter out noise and store an uncorrupted version of the signal. This works because changes to network connectivity add on time scales less than that of the decay. I now demonstrate this ability with an example where the signal corruption is due to both occlusion (part of pattern temporarily absent) and uniform noise (additional spurious inputs). I constructed a stimulus composed of two parts, signal and noise (Figure 2.6A). Different partially occluded versions of the pattern are presented briefly. Noise is also introduced, with other inputs randomly active such that the average firing rate is constant across all inputs.

In the context of such stimulation, it is not possible to distinguish between signal and noise by examining either any individual channel over all time, or all channels together at one individual point in time. However, because the plasticity integrates over all temporal associations on the second-long time scale, the noise ends up contributing much less to the connectivity compared with the more consistent signal over this time scale, resulting in an attractor dominated by the combinations of associates presented. By the end of training, presentation of a part of the pattern will activate a transient attractor corresponding to the entire pattern (Figure 2.6B), both filtering out the noise and filling in the majority of the occluded channels.

2.4.2 Modeling Attention and the Role of Inhibition

The transient attractor network also has the ability to turn its function on or off through straightforward modulation of inhibition. When the overall strength of inhibition is increased, recurrent activation of attractors will be suppressed such that the network will have no attractors (beyond faithfully relaying the stimulus). To demonstrate this, I take the network described in Figure 2.2, and rerun the simulations with the level of inhibition increased by doubling the strength of all inhibitory synapses. Although exposure to patterns still leads to synaptic strengthening, such changes are insufficient to create a stable attractor, and the final probe no longer leads to pattern recall (Figure 2.7). Such basic modulation coincides with observations of the requirement of attention or engagement for the storage of short-term

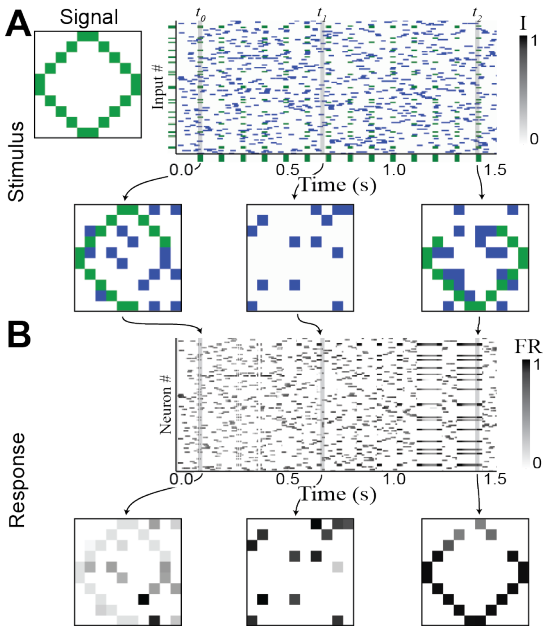

Figure 2.6: **Transient attractors able to de-noise and fill in occluded inputs. A.** The same transient attractor network in Fig. 2 is presented with input generated from partially occluded signal pattern (*top left*). The inputs corresponding to the pattern are labeled in green to the left of the vertical axis. However, the full pattern is never presented to the network, and instead 25% occluded signals (examples *bottom left* and *bottom right*) are presented at 0.1 sec intervals. In addition to signal, noise (blue) is presented in all non-signal inputs over the entire interval (e.g., *bottom middle*). Quantity of noise set so that each input has the same level of activity. **B.** Network activity in response to stimulus across all time (*top*) and in response to the three example inputs (below). Initially the network responds to noise and signal indiscriminately, but over time correlations in input allow it to filter out noise and fill in the occluded parts of the signal.

memories (D'Esposito and Postle, 2015), as well as for changes associated with auditory streaming (Shamma et al., 2013), and is generally useful to selectively perform the various functions of a transient attractor network.

2.4.3 Model Robustness

Stability is often a large concern in neural networks with recurrent excitation; a slight disturbance in the excitatory/inhibitory balance can either lead to runaway excitation or silence activity throughout the network. I tested how fine this balance is in the model by changing the baseline synaptic strengths (i.e., S_0 in eq. 2.1) of all neurons of a certain type, for example halving all feedback inhibition, and determining if the network continues to successfully store and recall patterns. Each individual parameter could be varied by at least 25% in either direction (Figure 2.8A), showing the model to be highly resilient to the average sizes of synaptic strengths. I attribute this stability to the close link between inhibition and excitation, as inhibition adjusts in line with general excitation, and the effect of saturating firing rates within the model.

Furthermore, while I have thus far assumed homogeneity in connections within the transient attractor network in order to maximize the number of patterns that could be robustly stored, the underlying functionality of the network is robust to significant variations of this homogeneity. To demonstrate this, I randomly removed a percentage of recurrent connections while keeping total recurrent connection strength

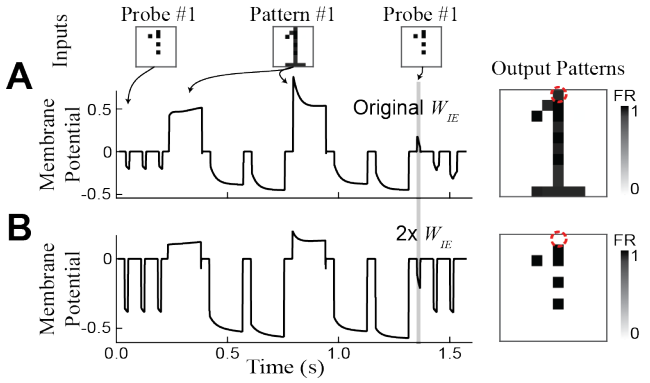

Figure 2.7: **Inhibition as proxy for attention.** The transient attractor network from Fig. 2 with varying levels of inhibitory feedback. **A.** Original levels of inhibitory feedback lead to sufficient increases in membrane potential at second probe (*left*), and hence complete pattern recall (*right*). The membrane potential is plotted for red-circled cell. **B.** Doubling the feedback from inhibitory to excitatory suppresses the membrane potential overall (*left*), and prevents memory recall (*right*), while still faithfully relaying the probe pattern.

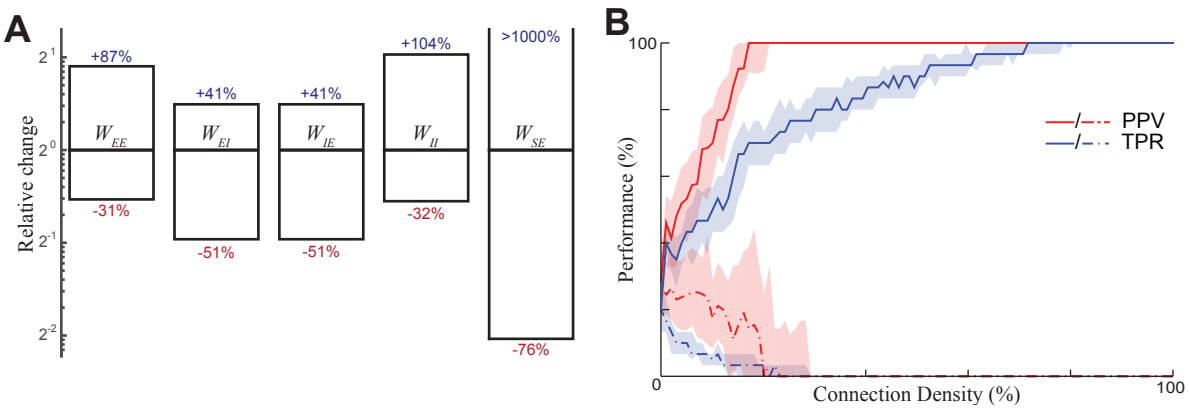

Figure 2.8: **Transient attractor model robustness to variations in parameters and network homogeneity.** Tests of parameter robustness are performed on the transient attractor network from Fig. 2, while storing two random non-overlapping patterns. **A.** Range over which each parameter may be varied without stopping successful storage/recall of information. Reported quantities in range are relative to default parameter values. **B.** The performance of the network in recalling both memories when the homogeneity condition is violated by randomly removing some fraction of recurrent excitatory connections. Performance is measured as Positive Predictive Value (PPV) and True Positive Rate (TPR) (see Methods). Dashed/solid lines are median before/after training, shaded region lies between first and third quartile.

constant. It was found that the network still functions remarkably well at recalling any pattern for connection densities as low as 20% (Figure 2.8B). This result comes from the manner in which memories are stored as associations between many different pairs of neurons which is only perturbed when a large proportion of connections have been removed. I conclude that the model is not highly dependent on the assumption of uniform connectivity.

The transient attractor model scales well to larger networks. In a larger network, the probability that two patterns significantly overlap significantly decreases (assuming either relative or absolute size of each pattern fixed), so memories are less likely to get confused. This is related to the fact that the memory capacity of a Hopfield network scales linearly with network size. Likewise, memories in larger networks are stored across multiple synapses, so that irregularities at any single synapse is less likely to cause issues.

2.5 Discussion

In this chapter I have presented the transient attractor network, defined primarily by recurrent excitatory connections that are governed by an associative (Hebbian) plasticity that only lasts on the scale of seconds. I have demonstrated that such a network is capable of a wide range of useful behaviors, including short-term memory (Figs. 2.1-2.3), source (or stream) segregation (Figure 2.4), signal de-noising (Figure 2.5), memory maintenance (Figure 2.6), and how these abilities can easily

be modulated by a simple top-down signal that modulates the level of inhibition (Figure 2.7). Furthermore, I demonstrated the robustness of the model with respect to both synapse strength and homogeneity (Figure 2.8). The concept that the same underlying network mechanism might have several uses in sensory computation is compelling in its simplicity. In fact, each of the tasks in Figures 2.2 and 2.4 - 2.7 was performed using the exact same network with the same parameters. Furthermore, while many of the above functions of transient attractor networks are demonstrated with these simplified networks, the networks size should actually make its desirable properties more robust.

The mechanisms and network structure underlying transient attractors are known to exist in the cortex except, perhaps, for associative transient plasticity (Section 2.5.2). It does not depend on a set of stable attractors, or some finely prescribed structure. This allows it to be a candidate for short-term memory in a wide variety of regions, such as the primary sensory cortex (Pasternak and Greenlee, 2005; Postle, 2015). This is in contrast with a large number of short-term memory models which prescribe such tasks to particularly specialized regions of the brain. The broad applicability of short-term memory benefits from widely applicable mechanisms, perhaps working in tandem with more specialized regions.

2.5.1 Alternative Models for Short-Term Memory

The classic model for short-term memory stores information in persistent attractors (Barak and Tsodyks, 2014), that is through a self-sustaining state within the network. Once such an attractor is activated, activity will persist until externally stopped, while the identity of the persistent attractor stores the information. This self-sustenance is typically achieved in neural networks through different combinations of recurrent excitation (Goldman, 2009), inhibition (McDougal, 2011), or both (Aksay et al., 2007; Lim and Goldman, 2013; Machens et al., 2005). Of the many models of persistent attractors, an interesting subset makes use of synaptic modifications to the attractor to aid in the persistence of activity (Barak and Tsodyks, 2007; Itskov et al., 2011). The combination of persistent activity and underlying synaptic modifications does resemble the transient attractor network (Figure 2.3D), but nevertheless information storage in these networks relies on persistent activity. While some experimental results appear to support the idea of persistent activity underlying short-term memories (Courtney et al., 1997; Fuster and Alexander, 1971; Seung, 1996), a number of recent studies in different brain areas have cast doubt on this general conclusion (Barak and Tsodyks, 2014; Mongillo et al., 2008; Stokes, 2015).

As a result, other models for short-term memory have been proposed, using processes such as cell assemblies (Lansner et al., 2003), non-stationary activity (Amit et al., 1997), cross-regional networks (Dubreuil and Brunel, 2016; Verduzco-Flores

et al., 2009), or purely feed-forward circuits (Goldman, 2009). These other ideas all rely on neural activity for information storage, and thus are still distinct from the idea of storing information in neural connectivity.

The concept of storing information within temporarily strengthened synapses goes back at least 25 years; for example, Shen (Shen, 1989) demonstrated its potential use for neural integrators. More recently, it has been shown how synaptic dynamics could be used to store short-term and temporarily inactive memories using either direct associative plasticity (Buhmann and Schulten, 1986; Sandberg et al., 2003; Szatmáry and Izhikevich, 2010) or synaptic facilitation (Mongillo et al., 2008). However, in all of these models the scope of the memories was pre-defined by the structure of the network: Sandberg et al. essentially used a ring attractor which could store individual variables due to the ring structure, Szatmary/Izhikevich used randomly created periodic attractors, while Mongillo et al. facilitated pre-defined cell assemblies. My work builds on this by incorporating the ideas of Dynamic Field Theory (DFT)(Schneegans and Schöner, 2008), that is considering how the attractor structure shifts in the presence of particular inputs. The most obvious ramification of this is the ability for targeted memory retrieval. More subtly, this expands the scope of memories that can be stored in the network; they no longer need to be stable attractors in their own right (as explored in the field of persistent attractors), or in the presence of background noise (in the three papers discussed above), but only during certain types of input. This in turn allows for novel points such as the completely flexible set of memories, the use of attractors in automatically processing

sensory inputs, their regulation by attention, or in their general robustness to noise.

2.5.2 Experimental Evidence for Transient Associative Synaptic Plasticity

The transient attractor network above relies on an associative learning rule that decays on the order of seconds. There is scattered experimental evidence for transient associative effects (i.e., where strengthening of connectivity occurs between coactive neurons), which has been observed in ferret auditory cortex (Shamma et al., 2013), macaque PFC (Sugase-Miyamoto et al., 2008), and dissociated networks (Dranias et al., 2013). It is known that associative learning takes place over a variety of timescales due to multiple mechanisms (Kandel, 2014), including some direct associative connections which decay in minutes (Erickson et al., 2011; Malenka, 1991). It is conceivable such processes might exist for shorter timescales, but have proven difficult to separate from non-associative plasticity similar timescales (such as synaptic facilitation and depression). Such associative plasticity also may be possible to achieve associative changes in effective coupling using non-associative facilitation within certain network structures; this is the subject of chapter 3.

2.5.3 Extensions of the Transient Attractor Network

It is hypothesized that the pre-existing wiring of neural networks in the sensory cortices is informed by the structure of natural stimuli (Hebb, 1949), which is equivalent to non-uniform connectivity (S_{ij}) in the transient attractor network. While

such non-uniformity would bias the network towards some attractors, this could be advantageous in the sensory cortices, as the location of transient attractors will be guided both by the immediate history and by the pre-learned nature of typical stimuli. When presented with a novel stimulus, the network's interpretation may be biased by learned stimuli, which are presumably the stimuli that have proven the most useful (given rules of long-term plasticity). This coordination of short- and long-term plasticity is distinct from earlier work that stored short-term memories by strengthening some pre-existing attractors: in the transient attractor model, recent activity may change the nature (e.g. strengthen, make stable or shift) pre-existing attractors. This allows for much greater flexibility in memory storage; the number of possible transient attractors (as influenced by pre-learned patterns, recent history, and by the nature of the instantaneous input) is far larger than that of pre-existing attractors.

Methods

Simulations

All simulations were performed in MATLAB using the Euler method with $\Delta t = 0.1$ ms.

Neuron Model

In all simulations I used a continuous firing rate model with cell voltage, $v_i(t)$, depending on the weighted sum of recurrent excitatory, inhibitory and input currents,

$$\frac{dv_i(t)}{dt} = -\tau v_i(t) + \sum_{Exc_j} W_{ij} y_j(t) + (E^I_{rev} - v_i(t)) \sum_{Inh_j} W_{ij} y_j(t) + In(t)$$

In this formulation, the excitatory input currents are independent of the cell membrane potential. This approximation is valid so long as the excitatory reversal potential is far higher than the firing threshold, which is the case for cortical neurons. In contrast, this can not be assumed for the inhibitory currents because the inhibitory reversal current is typically close to the cell membrane potential at rest. Firing rates, $y_i(t)$, are then calculated as a saturated rectified linear function of the cell voltage,

$$y_i(t) = \max[0, 1 - \exp(-a(v_i(t) - b))].$$

Parameters

Simulation parameters which remain constant across all simulations are listed in Table 2.1.

Weights between neurons depend on the network structures used in each Figure.

For Figure 2.1: $W_{SE} = 5$, $W_{IE} = 5$, $W_{II} = 20$, $W_{EI} = 10$, $W_{EE} = 1$

For Figure 2.3: $W_{SE} = 1$, $W_{IE} = 10$, $W_{EI} = 2$, $W_{EE} = 1.5$

For Figure 2.2, 2.4-9: $W_{SE} = 5$, $W_{IE} = 5$, $W_{II} = 20$, $W_{EI} = 1$, $W_{EE} = 0.1$

Ring Model

Network as portrayed in Figure 2.3. The recurrent excitatory baseline weights are Gaussian with mean = 1.5 (from Parameters) and standard deviation = 10. All weights are affected by a random multiplicative noise term, drawn from a normal distribution, $\mu = 1$, $\sigma = 0.05$.

Robustness analysis

Results in Figure 2.8. Parameter changes: Each parameter changed individually until memory recall is no longer successful. Recall is successful if, during a relevant probe, the average firing rate within either pattern is five times greater than maximum firing rate out of the pattern (including alternate pattern) and the average firing rate in each pattern is at least 0.1 (10% maximal firing rate).

Sparsity sensitivity: Connection density tested over range [0.05:0.05:1], 100 trials for each value. Each trial has a new, randomly selected connection matrix with each connection set to 0 with probability = density, and all remaining weights scaled uniformly to ensure total recurrent excitation remains constant. Each pattern contains 20 randomly selected excitatory neurons while each prompt is a subset of 5

Name	Symbol	Value
Max Hebbian	H_{max}	5
Min Hebbian	H_{min}	1
Hebbian increase	τ_{H_+}	100ms
Hebbian decrease	τ_{H_-}	2000ms
Depression increase	τ_{x_+}	50 ms
Depression decrease	τ_{x_-}	100 ms
Leak current	τ	0.5 ms^{-1}
Firing scale	a	1
Firing threshold	b	1
Inhibitory reversal potential	E^I_{rev}	-1

Table 2.1: **Transient Attractor Network Parameters** Above parameters were used in all simulations in the chapter.

from each pattern. An excitatory neuron is considered active if its average firing rate is over 0.1 while probe displayed. Analysis does not consider probe neurons in any category. PPV = number active within pattern / total number active, TPR = number active within pattern / number within pattern.

Chapter 3: Achieving Transient Associative Plasticity through Synaptic Facilitation

3.1 Overview

Associative ('Hebbian') synaptic plasticity acts to strengthen the connections between neurons which are active at the same time; this allows neural networks to store information for the duration of the plasticity. Such learning is of use to networks on a variety of time scales, including the ability to form transient associative connections which decay within seconds. However, experimental evidence for a direct type of this transient associative plasticity is lacking. Instead, typical mechanisms of short-term plasticity, such as synaptic facilitation, are not associative, meaning they are activated regardless of whether activity in one neuron is paired with its postsynaptic targets. In this chapter, I will demonstrate a network architecture by which facilitation could lead to short-term associative strengthening, and as a result be used to create a 'transient attractor' network capable of short term memory. In addition, it can perform temporal coherence analysis useful for auditory streaming, as well as robust stimulus de-noising. Thus, facilitation acting within specific cortical microcircuits could play a key role alongside other known synaptic effects to

control how information is processed throughout the cortex.

3.2 Introduction

The real world 'causes' of sensory inputs, such as objects generating visual or auditory stimuli, typically persist on time scales many times longer than those of activity from a single neuron. Making sense of natural sensory stimuli would therefore be greatly assisted by the ability to store information about the sensory inputs for some period of time. One common means to preserve information within a network for longer than individual neuron reaction times is in self-sustaining (or persistent) patterns of activity (Barak and Tsodyks, 2014; Seung, 1996). Recently, however, several experimental findings have raised doubt on whether such persistent activity is responsible for all the short term retention of information in the cortex (Postle, 2015; Sreenivasan et al., 2014). In this chapter, I will concentrate on an alternative mechanism for the storage of such short-term information: temporary modifications to the network's functional connectivity.

Changes to a network's connectivity will affect how the network responds to incoming signals; these changes might therefore allow for the storage and recall of short-term memory. One example of this is the transient attractor network presented in Chapter 2. This network is based on recurrent, excitatory connections that are strengthened due to coincident pre- and postsynaptic activity. Strengthening connections between coactive neurons means it is more likely they will be coactive in

the future. This strengthening is transient, and decays on the order of seconds. Such transient changes in connectivity modify the network's attractor structure: when a subset of neurons in a pattern are externally stimulated, excitatory currents to other neurons within the pattern are amplified by the strengthened connections. Later, the whole pattern may be recalled using an appropriate probe. Such transient attractors have been associated with cortical processes including short-term memory, temporal coherence analysis and signal de-noising.

The function of the transient attractor network relies on transient associative plasticity. Indeed, such modifications are known to occur directly on an individual neuron to neuron basis through mechanisms of long-term potentiation (LTP) (Tetzlaff et al., 2012). However, LTP is usually defined by changes occurring over time scales of minutes to hours. Although it may be possible that information that is only useful for seconds is saved for much longer periods of time due to such mechanisms, accumulating large quantities of obsolete information will make recalling individual pieces of information very difficult.

One such candidate mechanism for short-term plasticity on the time scale required for the transient attractor network is synaptic facilitation, whereby the strength of a synaptic input is temporarily strengthened due to presynaptic neural activity. This occurs due to a buildup of calcium in the presynaptic terminal which increases the probability of vesicle release (Zucker and Regehr, 2002). Such strengthening will occur regardless of whether it is coincident with postsynaptic activity. Such

non-associative plasticity cannot store memories in the all-to-all transient attractor network; any facilitation at one neurons synapses will affect all other neurons equally.

Sparser networks, however, may temporarily store information using non-associative short-term plasticity. This was demonstrated by Mongillo et al. (2008), who presented a network which contained several pre-determined cell assemblies, i.e. groups of cells with strong connections between cells in the same group. When a given assembly is active, all cells comprising it undergo facilitation. As a result, should one cell in the assembly fire soon afterwards, the facilitation will increase the post synaptic currents in all connected cells. This disproportionately affects cells within the same assembly. Mongillo demonstrated how this could lead to the spontaneous reactivation of any recently active assemblies in the presence of background noise, allowing for short-term memory via facilitation. The memories that may be stored in this network are, however, pre-defined by the set of assemblies: the network cannot arbitrarily associate any pair of features.

In this chapter, I will demonstrate how a neural network may achieve associative plasticity using only non-associative mechanisms (namely, synaptic facilitation). I start with a simple model which demonstrates how facilitation can lead to associative plasticity in networks with a predefined circuit motif. I will then apply this motif to a transient attractor network, demonstrating that it can store short-term memories. Finally, I will consider a large network in which the recurrent connections represent a set of more generalized features, and examine how facilitation in such

a network might allow for short-term memory via transient attractors. Together, these networks describe how short-term associative plasticity may be achieved using synaptic facilitation.

3.3 Results

In order to demonstrate how short-term associative plasticity might be achieved in a network of facilitating neurons, I start with a simple example. A small network is constructed which is capable of transient associative learning within a population of three 'basis' neurons without any form of associative plasticity. This means that stimulating any two basis neurons at the same time will lead to a temporary increase in recurrent excitation between those two neurons, but not to the third neuron. This was achieved through a sparse network structure; a diverse system of connections between neurons allows for non-associative mechanisms to have a targeted effect. In particular, the network includes a second population of neurons, 'feature' neurons. Each feature neuron connects to and from a pair of basis neurons, and one feature neuron exists for each pair of basis neurons (Figure 3.1A). In addition, the network contained one 'basis' inhibitory neuron and one 'feature' inhibitory neuron, to mediate competition within the two excitatory populations. All connections are initially equal in strength.

Synaptic strengths between excitatory neurons are the product of three factors: a constant, baseline weight S_{ij}, a synaptic facilitation factor $u_i(t)$, and a synaptic

depression factor $x_i(t)$. The latter two of these will be effected by the presynaptic firing rate, $y_i(t)$. The synaptic strengths are governed by the equations

$$W_{ij}(t) = S_{ij}u_i(t)x_i(t) \tag{3.1}$$

$$\frac{du_i(t)}{dt} = \frac{1}{\tau_{u_+}}\left(u_{max} - u_i(t)\right)\left(y_i(t)/y_0\right)^{\alpha} - \frac{1}{\tau_{u_-}}\left(u_i(t) - 1\right) \tag{3.2}$$

$$\frac{dx_i(t)}{dt} = \frac{1}{\tau_{x_+}}\left(1 - x_i(t)\right) - \frac{1}{\tau_{x_-}}x_i(t)u_i(t)y_i(t) \tag{3.3}$$

(Tsodyks et al., 1998). Here, I used common values of the time scales of both synaptic facilitation and depression decay, $\tau_{u_-} = 1.5s$, $\tau_{x_-} = 0.2s$ (Mongillo et al., 2008). Note that the dynamic terms are dependent only on presynaptic activity, and so will be applied equally to all synapses from the same neuron. In the presence of presynaptic activity, the synaptic facilitation will increase towards u_{max} and the synaptic depression $x_i(t)$ will decrease towards 0. In the absence of activity, both the facilitation and depression will asymptote to 1, with the depression recovering faster than excitation ($\tau_{u_-} > \tau_{x_-}$) (Hennig, 2013). This results in an immediate weakening of the synapse during activity, followed by a long period in which the synapse is stronger.

I next designed a stimulus to test associative learning between the three basis neurons. This will be composed of three periods. During the first, neuron B1 and B2 are activated; if associative learning occurs, this will lead to them being paired. In the second period, neuron B3 is activated, so that all neurons have been active an equal amount. This rules out excitability due to non-associative effects. I then tested for associative learning by stimulating one of the paired neurons (B2), and observing

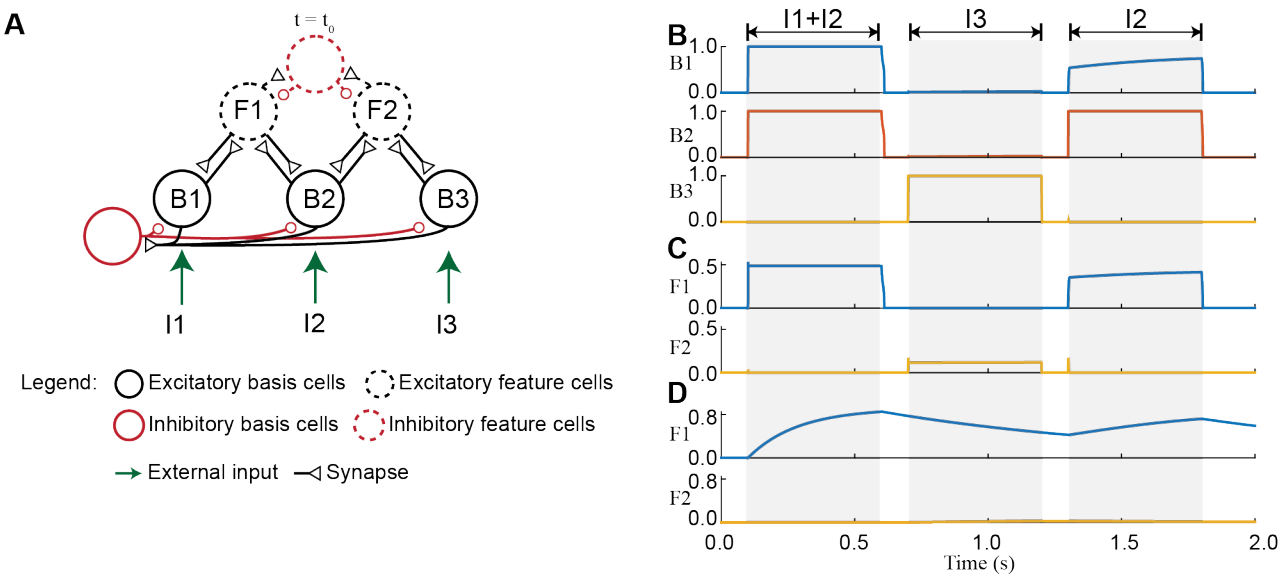

Figure 3.1: **A sketch of associative plasticity due to synaptic facilitation. A.** Network structure composed of three excitatory basis cells, three intervening feature cells (two shown), and a single (feature) inhibitory cell. External stimulus is designed to test Hebbian learning; the first two input neurons are initially co-active, while the third not co-active with any other neuron. **B.** Basis layer activity reveals Hebbian association has been achieved, as the network now exhibits (indirect) excitation between the neurons which were initially co-active, but not between those which weren't. **C/D.** Associative learning achieved through the level of activity and facilitation of the feature neurons. Feature corresponding to co-active pair of neuron shows much higher level of activity, and therefore facilitation. This in turn leads to more current flowing to the paired neuron (B1), providing further input current to the facilitated feature neuron (F1), and allowing the facilitated feature to outcompete the other feature neuron.

the resulting activation. During this final period, the paired neuron (B1) is active, whereas the unpaired neuron (B3) is not. This indicates associative learning has occurred. As this learning is controlled by facilitation, it will decay on the timescale of facilitation. Therefore, this demonstrates transient associative plasticity using non-associative mechanisms.

This process depends on facilitation ($u_i(t)$) of the feature neurons (Figure 3.1C). Feature neurons respond strongest, and hence facilitate strongly, when both of their basis neurons are active. This means that those feature neurons which represent previously co-active pairs of neurons will have strongly facilitated synapses, strengthening the connection between the neurons. Should one neuron in the pair become active, the other will also become significantly active, which will also increase the excitatory input into the feature neuron (positive feedback). In addition, the inhibitory neuron mediates competition between the feature neurons; those feature neurons which connect to one active and one inactive neuron may fire initially, but will soon lose out to other feature neurons which connect to two active neurons. This is also supported by synaptic depression; many postsynaptic currents may be initially strong enough to cause neural activity, but only those which then receive strong positive feedback (that is, they are associated with the dominant assembly) may remain active.

3.3.1 Short-Term Memory in a Complete Facilitating Network

Having established how transient associative plasticity can be achieved in our facilitating feature network, I next considered a specific application of transient associative plasticity: the transient attractor network. This network made extensive use of direct transient associative plasticity, and was shown to be capable of multiple types of cortical tasks within the sensory cortices. I now demonstrate how a facilitating feature network can mimic one of these tasks being the ability to store and recall multiple short-term memories.

I constructed a network based on the above schema, except with 9 basis cells (and consequently 36 feature cells), and test if two memories might be stored and recalled. A simple stimulus was designed for this task (Figure 3.2B). The majority of this stimulus involved imprinting the network with the patterns; the patterns were shown alternately at 4 Hz for 1 second. A probe was also selected for each pattern to test whether the pattern has been stored in the network. These probes are simply one neuron from each pattern. Each probe was shown briefly before and after the training period. Before training, the probe stimuli only led to the activation of the probe neurons themselves, indicating no memory is initially stored in the network. After training, however, each probe led to the recall of the entire pattern with which it is associated (Figure 3.2C). This transient attractor network is therefore capable of flexibly storing multiple patterns within its facilitated synapses.

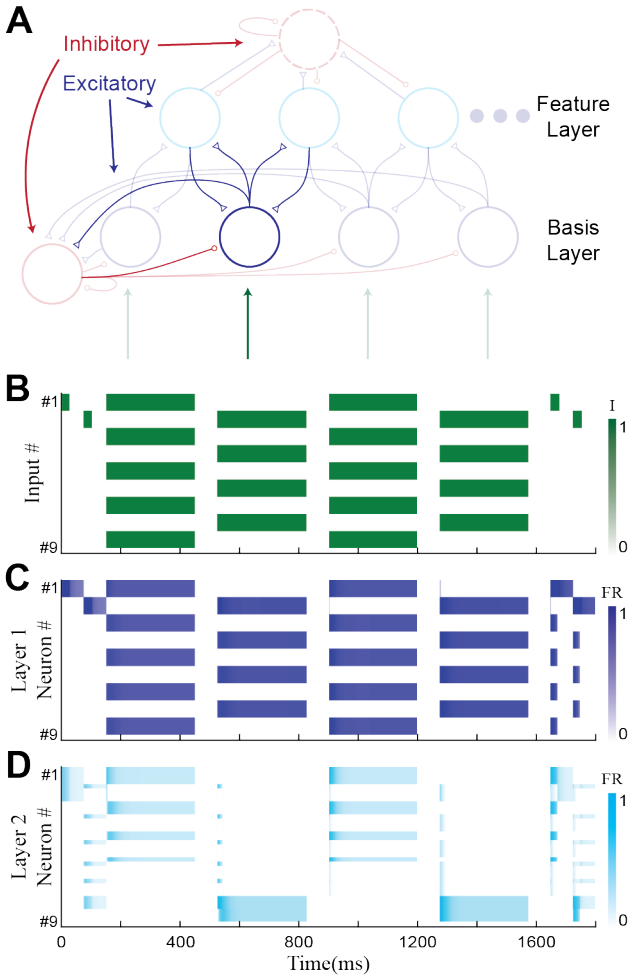

Figure 3.2: **Short-term memory in a two-layer, pairwise complete network. A.** Network structure with two layers with one neuron in second layer for each pair of neurons in first layer (subset of all second layer neurons shown). **C.** Stimulus composed of two presentations of each pattern (250 ms duration each, 50 ms between), capped at beginning and end by probe stimulus (subset of each pattern, 20 ms duration, 80 ms between). **D-E.** Activity of first (D) and second (E) layers of excitatory input in response to input. First layer successfully completes patterns after training.

3.3.2 Facilitating Feature Network with Generalized Features

The above transient attractor network contains simplistic features (a pair of neurons), with every possible feature represented. In this way, it can temporarily learn associations between any possible inputs. This is not necessary for processing natural stimuli which typically contain higher order correlations; not all pairs of inputs are equally likely to occur, or are as behaviorally important to the individual.

Instead, cortical networks are thought to successively decompose inputs into sets of increasingly complex features. The set of features chosen is not random, but instead works to efficiently represent natural stimuli (Section 1.4.1.1). This also means that the set of features is not complete; not every possible feature is included (many are of little use in processing natural stimuli). This motivates the design of a more realistic neural network, where the network is primarily concerned with storing information about some set of 'natural' patterns, as opposed to any pattern. It does this through the use of smaller, pre-learned features, which store information about those same patterns.

Once again, I constructed a network composed of basis and feature cells. I also randomly selected a set of P 'usual' patterns. These patterns will inform the selection of features, and the network will be tested to see how well it can store and recall one or two of these patterns. I do not assume that a stable attractor already exists for each distinct pattern (as would be required for persistent activity). Instead, re-

peated exposure to a pattern will cause facilitation of the relevant feature neurons, increasing disynaptic excitation between the basis neurons within the same pattern. This will then assist in later recalling the pattern.

The two-layer network contains 100 basis neurons (each with their own input), 500 feature neurons, and one inhibitory neuron for each layer (Figure 3.3A). Each feature represents a set of 5 basis neurons. Before forming connections between the layers I randomly selected P patterns, with each pattern consisting of 20 basis neurons. Each feature cell will represent a random subset of 5 basis neurons from a randomly assigned pattern. Note that only a very small proportion of all features are present in the network (0.1% of all features within a pattern when P = 50, or 0.0005% of all possible features).

As an illustrative example of how the network functions, I considered an example where the network is initially aware of 50 different patterns ($P = 50$). I then selected two of these patterns to be presented to the network, and checked if the network can store them in its facilitated synapses. Prompts were randomly chosen (subsets of 5 neurons from both pattern), and the stimulus was created by alternating the chosen patterns at 4 Hz for 1 sec, with prompts at start and end (Figure 3.3B, *top*). At the initial prompt, the network failed to complete the patterns (Figure 3.3B, *bottom left*), although some extra channels are active due to the choice of features. After training, however, the network successfully recalled both patterns (Figure 3.3B, *bottom right*). This demonstrates how facilitation has allowed for the accumulation of

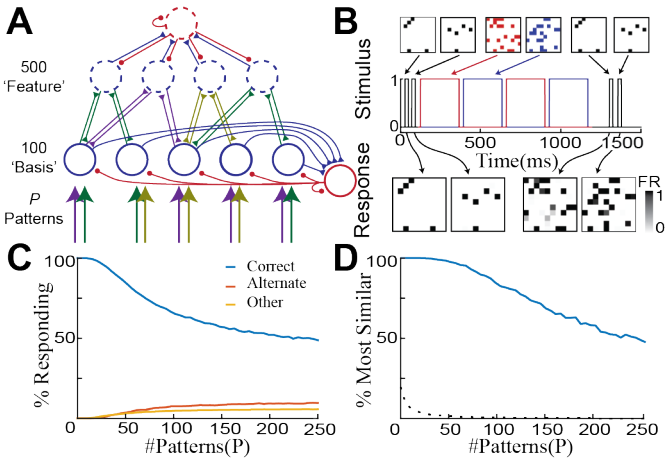

Figure 3.3: **Two-layer network with generalized features. A.** Network structure with wiring between layers determined by set of P randomly chosen patterns. **B.** Network successfully learns patterns. *Top:* Stimulus created using two patterns selected from set of $P = 10$ patterns, shown alternately (duration 250 ms, 50 ms between), capped at beginning and end by probe stimulus (subset of each pattern, 20 ms duration, 80 ms between). *Bottom:* First layer response before and after training showing successful recall of both patterns. **C.** Probability of firing for different classes of first excitatory layer neurons firing vs P, averaged over 1000 trials (prompt excluded). Blue line represents in pattern, red the alternate pattern chosen for training, orange all other neurons. **D.** Probability pattern most resembles correct pattern from set vs P, averaged over 1000 trials. Black dotted line is expected performance due to chance ($100\%/P$).

associations, creating transient attractors and allowing for the storage and recall of memories.

The degree to which the features in the network are tuned to represent each individual pattern depends on the number of recognizable patterns, P; an increase in pre-wired patterns in the network will mean fewer second layer neurons assigned to each pattern and more distractions from other feature neurons. For these reasons I anticipated a drop in performance with an increase in the number of recognizable patterns. I measured this performance in two ways.

First I examined which neurons are active in response to each prompt (Figure 3.3C). Below $P = 50$ the network recalls all patterns with very high accuracy, but even out to $P = 250$ the basis neurons within the pattern remain several times more likely to respond than the other neurons. To determine whether this indicates the network is recalling the correct pattern, or just a similar one, I measured how often the output is most similar to the desired pattern (compared with $P - 1$ other patterns used to determine connections). This probability decays with increasing P, so that the network's response has a 50% chance of most resembling the desired pattern when a total set of 250 patterns is used to determine feature set (Figure 3.3D). I remark that the network still performs remarkably well for high P values, when only a small number of second layer neurons are allocated to each pattern. In this case, successfully storing and recalling patterns involves incorporating features from other similar patterns without confusing the patterns entirely. I conclude that this network can store and recall a wide variety of patterns.

3.3.3 Extracting Information using Temporal Coherence

The transient attractor network (Chapter 2) is also able to segregate simultaneously presented patterns whose amplitude varies in time via temporal coherence (Shamma et al., 2011), and this property extends to the network presented here. Specifically, two patterns were chosen, and each pattern's amplitude over time was governed by a temporal envelope. I now demonstrate how this property extends to the network

presented here.

I used the larger transient attractor network described above, with wiring determined by a set of $P = 50$ patterns. Two patterns are selected from this set, and prompts chosen as subsets of each pattern. For training, each pattern is allocated a randomly varying temporal envelope (Figure 3.4A). In order to separate out the patterns, the network must learn which inputs share a temporal envelope. By comparing the network's responses to the initial and final prompts, it can be seen that the network can store, and recall, each individual pattern (Figure 3.4B). I conclude that this network is capable of performing some on-line temporal coherence analysis, and complete a simple streaming task.

3.3.4 Signal De-noising

By selectively strengthening connections within the networks, transient attractors may also act as integrators, and in doing so work to de-noise the input signal. Previously this was done using Hebbian synaptic learning; here I show it can also be achieved using facilitating synapses. (Figure 3.5A). This stimulus incorporates occlusion (signal partially missing) and noise in a way so that it is not possible to extract the signal using only information from any one point in time, or information from any single input channel. The stimulus used is the same as that used in chapter 2 on signal de-nosing.

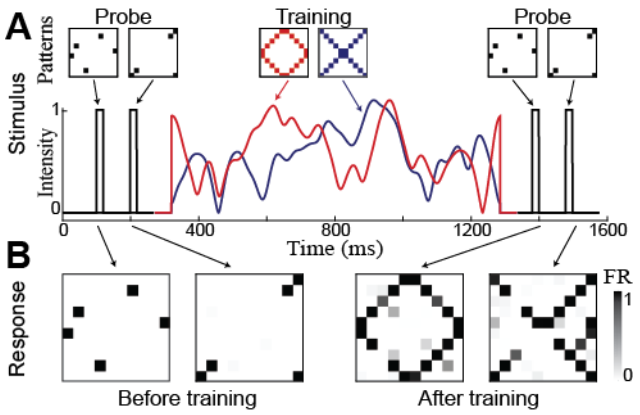

Figure 3.4: **Temporal coherence analysis with facilitating network.** From simulation with 100 first layer excitatory neurons, 500 second layer excitatory neurons and $P = 50$. **A.** Stimulus structure. Two patterns (*top, center*) are each assigned their own random temporal kernel (bottom, center). A subset from each pattern is chosen as a prompt (*top, left/right*), displayed briefly at beginning/end to test memories stored in network (*bottom, left/right*). **B.** Activation of first layer of excitatory neurons before (*left*) and after (*right*) training reveal separations of stimulus into two patterns using temporal coherence, with patterns recalled via suitable prompt.

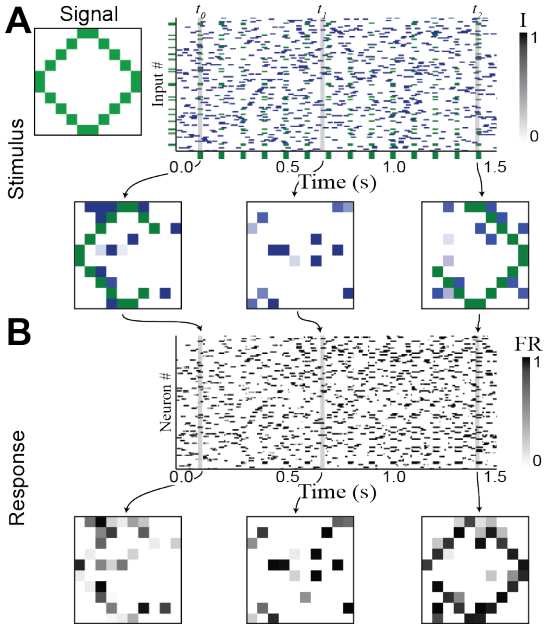

Figure 3.5: **Signal de-noising with facilitating network**. From simulation with 100 first layer excitatory neurons, 500 second layer excitatory neurons and $P = 50$. **A.** Stimulus composed of two parts. *Top:* Occluded patterns (25% occluded at a time) shows for 25 ms, repeats every 75 ms. *Bottom:* Additive noise random across all non-pattern channels, designed so all channels have approximately equal average firing rates and temporal duration. *Right:* The total stimulus at two points. **B.** Network activity in response to stimulus. Initially network responds to noise and signal equally, but over time correlations in input allow it to filter out noise and complete the pattern.

Initially, the network structure failed to discriminate signal from noise; the biased wiring structure does meant a non-trivial attractor was seen (Figure 3.5B). By the end, however, the network effectively distinguished noise from signal, and even included many parts of the signal which were temporarily occluded. Note that this is not the same as merely matching the stimulus with some attractor set (if so, the initial presentation would have led to a successful recall of the pattern). Instead, the network accumulated information about the likely nature of the signal by facilitating relevant synapses, and used this to usefully process the signal.

3.4 Discussion

In this chapter, I have demonstrated how short-term facilitation might mimic transient associative plasticity. Examples in this chapter are largely concerned with the transient attractor network introduced in chapter 2, which was shown to perform a variety of tasks related to sensory processing. These findings might prove to be applicable to several other neural processes. Associative plasticity allows neural networks to directly track correlations between neurons, and is widely associated with cortical processing. Direct associative plasticity has been observed to occur on many different timescales, but typically enduring at least tens of minutes (Tetzlaff et al., 2012). In contrast, non-associative mechanisms are known to endure on a variety of shorter timescales, from seconds up to a few minutes (Fisher et al., 1997). The above work might therefore allow those cortical processes which depend on associative plasticity to be performed on any time scales, including the timescales on

which stimulus objects typically persist (hundreds of milliseconds to tens of seconds).

Additionally, several models have already been proposed which rely on transient associative plasticity which decays within seconds. This includes models of short-term memory (Sandberg et al., 2003; Szatmáry and Izhikevich, 2010), auditory streaming (Von der Malsburg and Schneider, 1986), corrective prediction (Schultz and Dickinson, 2000) and in control of persistent activity (Brunel, 2003). The above concept of transient associative plasticity via short-term facilitation might allow for more realistic implementations of many of these models.

The short term storage of information in neural network through facilitation is only possible in asymmetric (e.g. sparse) networks; biased connections permit biased effects from facilitation. The use of sparse networks in this way was first exploited by Mongillo et al. (2008). In the model presented by Mongillo et al., only objects which correspond to pre-existing cell assemblies may be stored. The network presented in this chapter extended this idea by including a subpopulation of feature neurons; memories are then stored by facilitating a subset of all features, from which the object may be recreated.

This framework relies on a particular network structure within cortex. For one, it distinguishes between neurons as either basis or feature cells; such a division is not known to exist in cortex. Nevertheless, it is possible that a similar mechanism could allow for the short-term retention of information in a network without a clear

division between basis and feature cells. For example, when two cells are co-active, their shared neighbors are disproportionately likely to be coactive, and may therefore facilitate. Later, should one cell become active, it will disproportionately affect the other cell by way of the various facilitated, shared neighbors. The framework also assumes that recurrent connections between excitatory neurons are largely symmetric; there is some experimental evidence to support this (Ko et al., 2011; Song et al., 2005).

This chapter presented two different structures for facilitating networks. The first assumed that every possible feature is present; this may be of use in some small neural networks, but is inefficient for large networks. I then extended this idea to a more general concept of associative plasticity: when only a small subset of all possible features are represented by feature cells, facilitation may allow for recurrent connectivity between neurons associated with a recently seen feature. In biological neural networks, it is widely believed that successive layers of stimulus processing do indeed decompose stimuli into successive features (Section 1.4.1.1), and it is quite plausible that many of these connections would be bidirectional. Moreover, due to shared higher order correlations in natural stimuli, the number of features used to represent natural stimuli is far smaller than the set of all possible features. This means that the above facilitating transient attractor network makes use of a network structure that has been well documented within the sensory cortices to temporarily store information about any natural stimuli to which it was recently exposed.

Methods

Simulations

All simulations were performed in MATLAB using the Euler method with $\Delta t = 0.1$ ms.

Neuron Model

In all simulations I used a continuous firing rate model with cell voltage, $v_i(t)$, depending on the weighted sum of recurrent excitatory, inhibitory and input currents,

$$\frac{dv_i(t)}{dt} = -\tau v_i(t) + \sum_{Exc_j} W_{ij} y_j(t) + (E^I_{rev} - v_i(t)) \sum_{Inh_j} W_{ij} y_j(t) + In(t)$$

In this formulation, the excitatory input currents are independent of the cell membrane potential. This approximation is valid so long as the excitatory reversal potential is far higher than the firing threshold, which is the case for cortical neurons. In contrast, this can not be assumed for the inhibitory currents because the inhibitory reversal current is typically close to the cell membrane potential at rest. Firing rates, $y_i(t)$, are then calculated as a saturated rectified linear function of the cell voltage,

$$y_i(t) = \max[0, 1 - \exp(-a(v_i(t) - b))].$$

Parameters

Simulation parameters which remain constant across all simulations are listed in Table 3.1.

Weights between neurons depend on the network structures used in each Figure.

For Figure 3.1-2: $W_{SE1} = 10$, $W_{E1E2} = 20$, $W_{E2E1} = 20$, $W_{EXIX} = 1$, $W_{IXEX} = 20$, $W_{IXIX} = 10$, $y_0 = 5$

For Figure 3.3-5: $W_{SE1} = 10$, $W_{E1E2} = 5$, $W_{E2E1} = 5$, $W_{EXIX} = 1$, $W_{IXEX} = 10$, $W_{IXIX} = 10$, $y_0 = 5$

Name	Symbol	Value
Max Facilitation	u_{max}	5
Facilitation increase	τ_{u_+}	100ms
Facilitation decrease	τ_{u_-}	1500ms
Depression increase	τ_{x_+}	50 ms
Depression decrease	τ_{x_-}	100 ms
Leak current	τ	0.5 ms^{-1}
Firing scale	a	1
Firing threshold	b	0
Inhibitory reversal potential	E^I_{rev}	-1

Table 3.1: **Facilitating Network Parameters** Above parameters were used in all simulations in the chapter.

Chapter 4: Auditory Streaming via Gamma Partitioning

4.1 Overview

Understanding the auditory landscape around us involves discriminating between several auditory sources. The cortical processes that allow for this are not well understood. In this chapter, I will investigate how a biologically realistic network might represent multiple perceptual objects through a process of gamma partitioning. This partitioning incorporates both segmentation, that is the division of the the stimulus into multiple objects, and segregation, that is the separation the cortical representation of each object. I demonstrate how this online mechanism might be applied in processing several different types of auditory stimuli. I finish by considering how a biologically plausible online clustering technique might allow for the formation of coherent streams (so-called integration). This method is of particular interest due to its ability to exploit multiple features in forming streams, giving a plausible mechanism behind streaming in the auditory cortex.

4.2 Introduction

The auditory landscape that surrounds us is highly complex. Consider, for example, a cocktail party with music, multiple voices, and a variety of background noises. The sound we hear is a single waveform, the sum of individual sound waves from each source. 'Auditory streaming' refers to how the auditory system separates this combined information into auditory objects which represent the sources of the sound. This process has been divided into three parts (Bregman, 1990): segmentation, segregation, and integration (defined in Section 1.7).

Understanding the challenges involved in auditory streaming can be assisted by comparing it with visual object recognition. Distinct visual objects typically appear in separate regions of the two dimensional visual field, stimulating different regions of the primary visual cortex and therefore allowing for spatial segmentation and segregation. In contrast, sounds are primarily represented over a single dimension, frequency. When decomposed into their constituent frequencies, the sounds from individual sources are not continuous and localized, but disjoint and span a large range of frequencies. This means that the representation of different auditory sources significantly overlap. Auditory streaming is influenced by some non-spectral features such as spatial location, this information is not necessary for streaming; this is evidenced by the ability to stream sounds using a single audio channel.

The auditory system employs a large feature expansion when interpreting incoming

sounds; there are approximately 8,000 times more neurons in the auditory cortex than receptors in the cochlea (Worden, 1971). These neurons are collectively sensitive to a wide array of characteristics (Section 1.6.2). It is also known that sounds which are more similar with respect to these characteristics are more likely to be perceived as a single source (Section 1.6.3.2). We are left, however, with the so-called binding problem (Section 1.7.2): how do you represent the sets of features which are bound into a single stream?

A solution to the binding problem is suggested by correlation theory (Von der Malsburg, 1981): features that are represented at the same points in time are 'bound' into a single object. This theory has been employed in neural network models to explain auditory streaming (Von der Malsburg and Schneider, 1986; Wang, 1996) via targeted recurrent excitatory connections (Section 1.7). These models use networks with simple, local connectivity, concentrate on non-overlapping objects, and are composed of idealized 'oscillator units' as opposed to true neurons. These models will be further compared with the model presented here in the Discussion.

In this chapter, I will investigate a new model for auditory segmentation and segregation based on synchrony of neurons in the auditory cortex during gamma wave oscillations. These are oscillations that occur in cortex at between 30 Hz and 100 Hz (Arnal and Giraud, 2012) (Section 1.4.2.2). Gamma partitioning of sensory stimuli has previously been used to explain visual processing (Miconi and Vanrullen, 2010). This model used localized recurrent connections (local using a combination of loca-

tion and orientation). Auditory segmentation provides a much more challenging task due to the low dimensionality of the input, and the manner with which the auditory sources significantly overlap in their frequency representation. In this chapter, I will demonstrate how gamma partitioning can process auditory stimuli with minimal delay using non-localized recurrent connections. I will start by describing the model structure, then use it to process a variety of auditory stimuli. I also consider how this partitioning may lead to complete auditory streaming by combining it with a clustering algorithm over successive windows of time.

4.3 Results

A neural network is created which is composed of spiking, noisy, leaky integrate-and-fire neurons with 2 ms absolute refractory periods. Within the network there are three populations of neurons: one excitatory population and two inhibitory populations. The excitatory neurons are responsible for receiving and interpreting the stimulus, while the two inhibitory populations respectively mediate local and global competition within the excitatory population.

Groups of excitatory neurons are arranged tonotopically over the range 50 Hz to 2 kHz, spaced every 5 Hz. Each group contains multiple (100) spiking neurons; this allows us to smoothen the firing rate over time while maintaining realistic spiking behavior. The input current to each group is calculated using a Difference of Gaussians weight scheme (center σ = 10 Hz, surround σ = 100 Hz). There are also

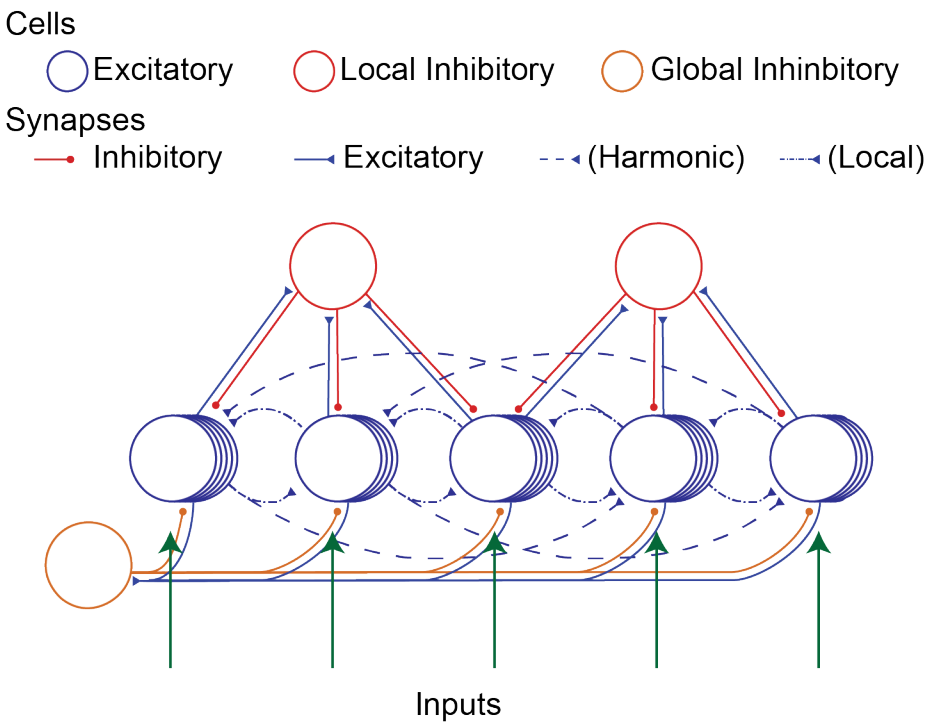

Figure 4.1: **Network structure.** Network composed of excitatory neurons, global inhibitory neurons and local inhibitory neurons. Recurrent excitatory connections change between different networks investigated.

recurrent excitatory connections between the groups of excitatory neurons; these inform the grouping of frequencies into objects, and are further discussed below.

Local inhibitory neurons enforce competition within groups of excitatory neurons, and between groups which represent similar frequencies. These local inhibitory neurons are arranged tonotopically over the range 50 Hz to 2 kHz, spaced every 20 Hz. They take inputs from, and project back to, nearby excitatory neurons; in my simulations, the connections are described by a Gaussian function centered on the neurons assigned frequency with standard deviation of 100 Hz.

The global inhibitory neuron connects equally to and from all excitatory neurons in the network. By integrating over all excitatory cells, this global inhibitory neuron monitors the total level of network activity, inhibiting all neurons if some threshold is passed. This behavior ensures that only a minority of cells may be active in any small window of time: in a sense, the global inhibitory neuron acts as a gatekeeper. The existence of this cell leads to gamma waves within the network - activity slowly builds, the global inhibitory circuit is activated, and all activity is subdued before starting the cycle once again.

4.3.1 Recurrent Excitatory Connections

This network may contain recurrent connections between excitatory neurons. These connections inform which neurons will be co-active; neurons which are connected are more likely to fire during the same gamma wave. All connections are symmetric in this network. I consider three different patterns of recurrent wiring: no recurrent connectivity, local recurrent connectivity, and harmonic recurrent connectivity.

These recurrent connections are determined using a Hebbian learning scheme (except for the case with no recurrent connections). This means that the connection strength is determined by average co-activity given some set of 'typical sounds (both typical 'local' sounds to determine the local recurrent connectivity and typical 'harmonic' sounds for the harmonic connectivity). This involved first constructing sets of such 'typical sounds', each sound has a random fundamental frequency between 250 Hz

and 1 kHz. Such 'local' sounds only contain energy at the fundamental frequency, whereas 'harmonic' sounds have energy at all integer multiples of the fundamental frequency. These energy distributions are then smoothed by convolving with a Gaussian function (σ =10 Hz). The recurrent excitatory weights are then calculated as

$$W_{ij} = [1 - \exp(-m(\rho_{ij} - \rho_{min}))] \tag{4.1}$$

where the correlation (ρ_{ij}) is calculated using a set of 10^6 'typical' sounds, $m = 5$ and $\rho_{min} = 0.1$. Under this training system, the harmonic connections will also include connections between neurons which represent similar frequencies.

When combined with the local inhibitory network, both the local and harmonic recurrent connections form a center-surround feedback pattern between the excitatory neurons - activity from an excitatory neuron excites its closest neighbors, but indirectly inhibits more distant neighbors.

4.3.2 Stimulus Pre-Processing

Before being fed into the three models, any auditory stimulus is pre-processed in a manner which is broadly faithful to cochlear processing (Figure 4.2). The cochlea is known to decompose the input waveform, measuring the amount of energy at each frequency; I mimic this using a discrete fourier transform, which is then smoothed using the aforementioned 'Difference of Gaussians' filter.

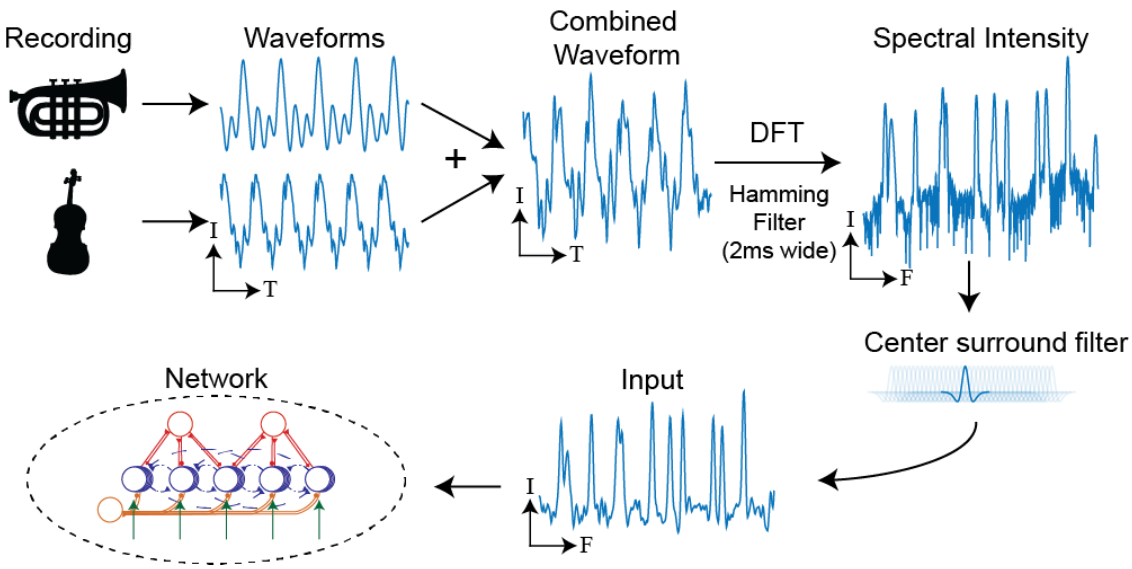

Figure 4.2: **Cochlear pre-processing.** Stages involved in pre-processing stimulus, similar to calculations performed in the cochlea. Sound waves from multiple sources are combined into a single waveform, which is then decomposed into frequency space using a discrete fourier transform (using a hamming filter 2 ms wide, phase information discarded). This is then passed through a 'Difference of Gaussians' filter to create the input current for each excitatory group of neurons.

4.3.3 Stimuli from Musical Instruments

I start by considering how each model responds when presented with the relatively static harmonic sounds produced by musical instruments. Recordings chosen were recorded by the London Philharmonic Orchestra (London Philharmonia, 2016). Recordings from four instruments (trombone, cello, clarinet, flute) were used that represent the main categories of non-percussive instruments (brass, string, reed and non-reed woodwind) and for which the database was most complete. Inputs are generated using 12 notes that form an octave within the range of all four instruments (C4 to B4 in International Pitch Notation, also known as the octave directly above 'Middle C'). This makes for a total of 48 samples with fundamental frequen-

cies between 250 Hz and 500 Hz.

An auditory stimulus is created by choosing two samples. I start with an in-depth analysis of the networks' behavior in response to one particular auditory stimulus (Cello C4 and Flute D4) (Fig 4.3A). There are several similarities in neural activity across all models: all exhibit gamma wave dynamics of similar frequency and intensity, and in all cases activity is concentrated in frequencies associated with inputs. There are also significant qualitative differences between the networks' behavior. The causes of these can be understood by examining the recurrent connectivity. Neurons (which represent individual bands of frequency) which are strongly coupled are highly likely to be simultaneously represented. This means that local connections cause neurons representing similar frequencies to co-activate, and harmonic connections cause entire harmonic stacks to co-activate. These harmonic stacks match well with the harmonic input from the individual sources.

I next established two different measures to quantify this network behavior. The first involves processing the network activity to separate it into different gamma waves, and then measuring how well each gamma wave correlates with the two different sources (over the interval in which the gamma wave occurred). These correlations may be plotted against one another, and the distance from the diagonal then used to estimate how much each gamma wave 'specializes' in either source (Figure 4.4A). These results confirm the above qualitative findings. In particular, individual gamma waves in the harmonic network represent individual sources with

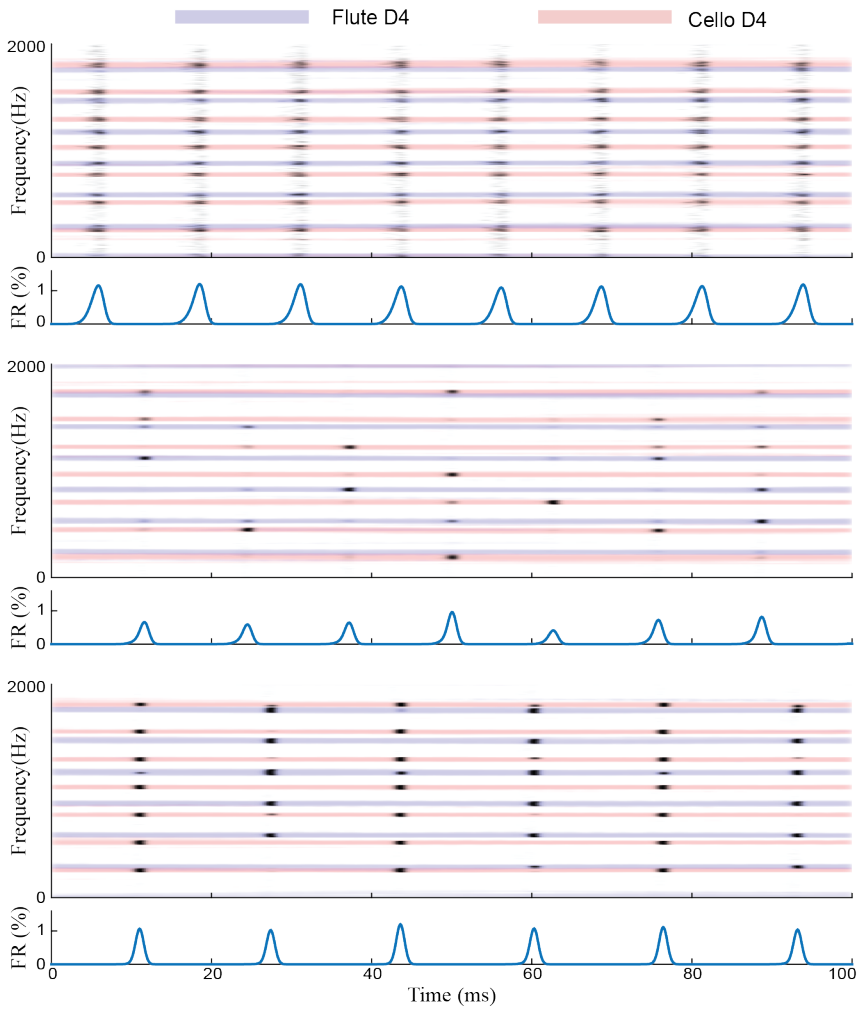

Figure 4.3: **Gamma partitioning applied to two instruments.** Network activity for three different patterns of recurrent connectivity (representative 100 ms shown of total 3 second simulation, activity smoothed over 1 ms to increase visibility). Network without any recurrent connections (*top*) has activity in every gamma cycle distributed uniformly across all harmonics. In contrast, each gamma wave in the network with local recurrent connections (*middle*) is concentrated in relatively few harmonics, but with no clear pattern between gamma waves. Finally, in the network with harmonic recurrent connections (*bottom*), each gamma wave matches fairly well with one source or the other, with the source represented alternating between subsequent cycles.

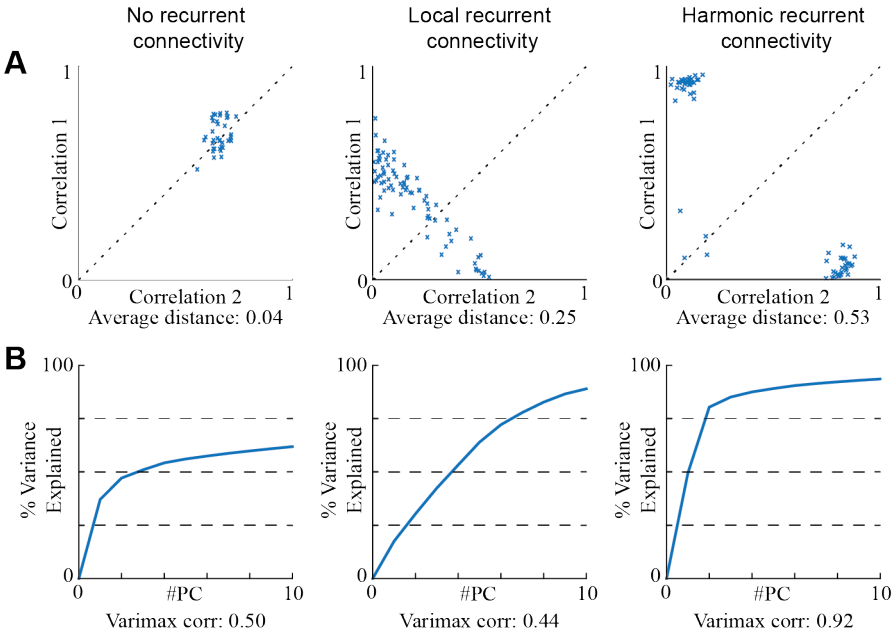

Figure 4.4: **Quantifying source separation.** Analysis of stimulus displayed in Figure 4.3. **A** Absolute value of correlation between individual gamma waves and two sources. Average distance from diagonal below. Note that the optimal score is $1/\sqrt{2} \approx 0.7$. *Left:* No recurrent connectivity shows all gamma waves representing both sources equally. *Middle:* Local connections cause significant variation between gamma waves, but none significantly match with either source. *Right:* Harmonic connections cause all gamma waves to correlate very well with one source, but not with the other. **B.** Principal component analysis and correlation between Varimax rotations of first two PCs and sources. *Left:* No recurrent connectivity has one significant PC, and varimax components do not correlate well with sources. *Middle:* Local connections cause several important PCs which collectively explain most network behavior, and first two PCs (rotated) do not match well with sources. *Right:* Harmonic connections leads to 2 highly significant PCs, which when rotated match well with two sources.

a very high level of fidelity.

The second measure attempts to test for a more general 'perceptual object'. Synchronicity theory need not dictate that only one object is represented in each gamma wave as the above analysis would suggest, but rather that the features attributed to each object follow a common temporal pattern. We tested for such general 'perceptual objects' using Principal Component Analysis (PCA); the number of principal components (PCs) required to explain most of the variance can measure the number of objects perceived. Analysis using PCA agrees with above findings (Figure 4.4B); no recurrent connections leads to a single object, local connections lead to several objects (the number of which is approximately equal to the number of harmonics present), and harmonic connections lead to two perceptual objects. Finally, we attempt to determine how these detected objects might match with the underlying sources. We do this using a Varimax rotation (Kaiser, 1958) on the first two PCs. The Varimax rotation is an orthogonal rotation which maximizes sparsity in individual components. The sounds from natural sources are often sparse in frequency, meaning that applying a Varimax rotation might allow the extraction of objects which are similar to stimuli. This was indeed the case for the harmonic network.

Finally, all of these metrics were applied for all possible pairings of instrument stimuli (excluding those in which both were playing the same note) (Figure 4.5). These results suggest that the above findings hold for just about all possible pairings of instrumental stimuli; no recurrent connectivity leads to a single perceptual object

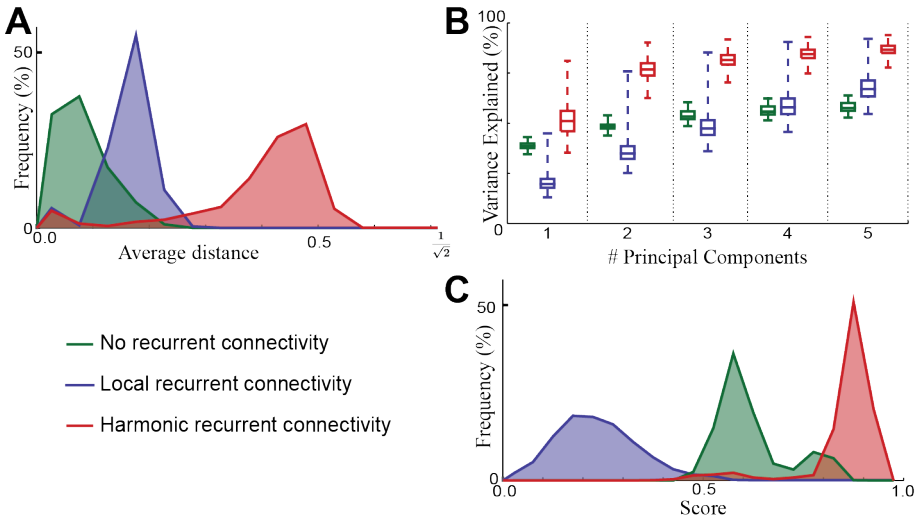

Figure 4.5: **Gamma partitioning across all samples. A.** Degree of 'specialization' as in Figure 4.4A. **B.** Number of 'perceptual objects' using PCA. **C.** How well first two PCs (rotated) match with two input sources.

which slightly correlates with both sources, local recurrent connectivity leads to multiple objects which don't correlate well with the sources, and harmonic recurrent connectivity leads to the detection of two perceptual objects which match well with the input sources. This suggests that gamma partitioning may separate many different combinations of harmonic stimuli.

4.3.4 Dynamic Vocal Stimuli

Thus far it has been shown how gamma partitioning might work to disentangle combinations of harmonic sounds in a network with harmonic recurrent connectivity. To simplify analysis, I have only used stimuli which have an approximately constant harmonic structure. The majority of auditory stimuli, however, change significantly on the time scale of hundreds of milliseconds or less. We therefore now consider how the auditory partitioning provided by the network with harmonic recurrent connec-

tivity might assist with the streaming of such dynamic sounds.

To test this method, I have chosen the first four bars of Queen's 'Bohemian Rhapsody' (Mercury, F, 1975). This song was selected due to the existence of a multitrack recording with multiple vocal parts, not as an ill-conceived attempt to win favor with my thesis committee. Two distinct vocal parts were selected, combined into a single waveform, and any silent periods were removed. Finally, the frequency was shifted slightly so that the whole sample would fall between 250 Hz and 500 Hz; this was done by rescaling time, speeding up by 40%.

Due to the constantly evolving nature of both the underlying sources and the perceptual segments, we use the correlation between each gamma wave and each (instantaneous) source to determine how gamma partitioning has processed the stimulus (Figure 4.6). As before, it can be seen that most gamma waves correlate strongly with one of the sources (as they existed at the moment of each gamma wave). This is a strong indication that gamma partitioning can successfully handle dynamic stimuli.

It is interesting to ask how easily these segments can be joined together into coherent auditory streams. This can be done using the continuous representation hypothesis (Section 1.7.3); each segment is paired with the preceding segments with which it was most similar. To do this we turn to the literature for the online clustering of data streams (Venkatasubramanian, 2009). A sliding k-means clustering (k=2) al-

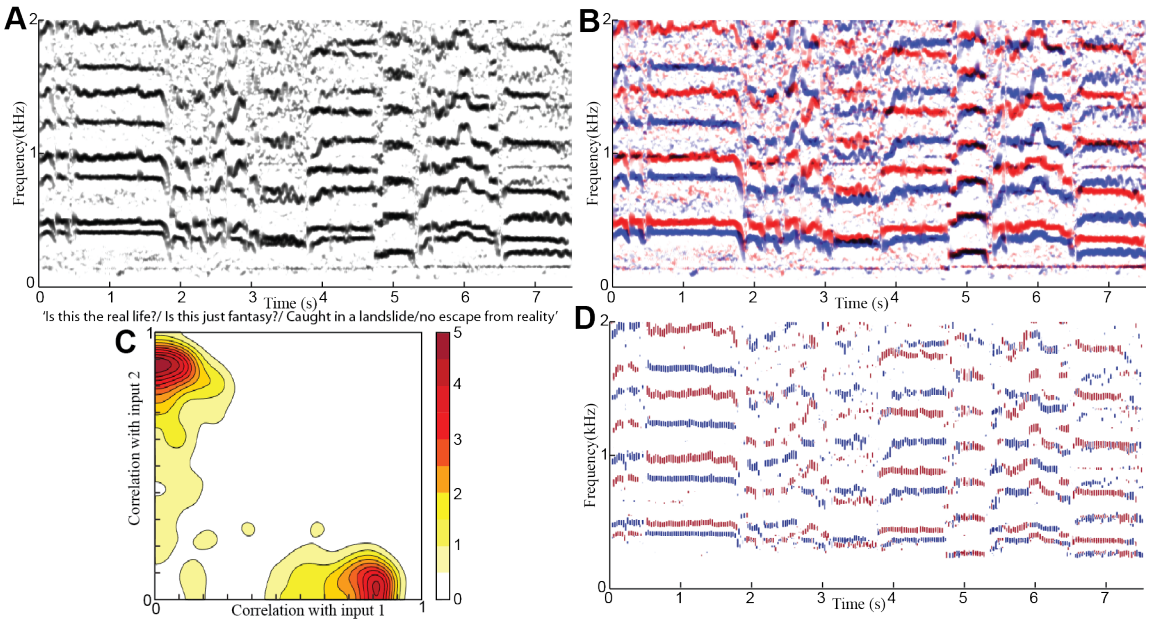

Figure 4.6: **Gamma partitioning of 'Bohemian Rhapsody'. A.** Input stimulus composed of two vocal parts of introduction. **B.** Parts labelled using individual recordings **C.** Correlation between individual gamma waves and each individual source reveals most gamma waves bear a strong resemblance to one of the sources. Average distance from diagonal = 0.57. Smoothed heat map (reflected at boundaries) displayed rather than raw data points due to quantity of data points. **D.** Auditory streaming can be performed by combining gamma partitioning with the sliding k-means algorithm to join segments together. Stream identities swap when sounds are discontinuous or significantly overlapping.

gorithm was applied both due to its simplicity and because k-means clustering can be performed by some forms of neural networks (Murtagh and Hernández-Pajares, 1995). Windows are 50 ms long (neurally plausible timescales), neighboring windows overlap by 90%, and clustering is performed on timepoints with a significant amount of neural activity (>0.01% average). Stream continuity was achieved by initializing each cluster center with those found in the previous window.

This complete streaming algorithm shows some success at forming coherent streams (Figure 4.6D). This algorithm does encounter issues when the underlying auditory objects are discontinuous in frequency space, or when they significantly overlap one another. In theory, this could be remedied by including a wider variety of auditory features (influenced by factors such as timbre or spatial location) which change slowly in time.

One final example of auditory streaming via gamma partitioning has been included (Figure 4.3.4). This was done on the opening phrase of the 'Flower Duet from Delibes' Lakmé (Delibes, date unknown). Unfortunately, I was not able to obtain individual recordings of each singers part, and so I was not able to quantitatively confirm the network performance. However, it is possible to see that the network segregates the input into individual streams, and that all harmonics within each stream undergo similar modulations in frequency (trills).

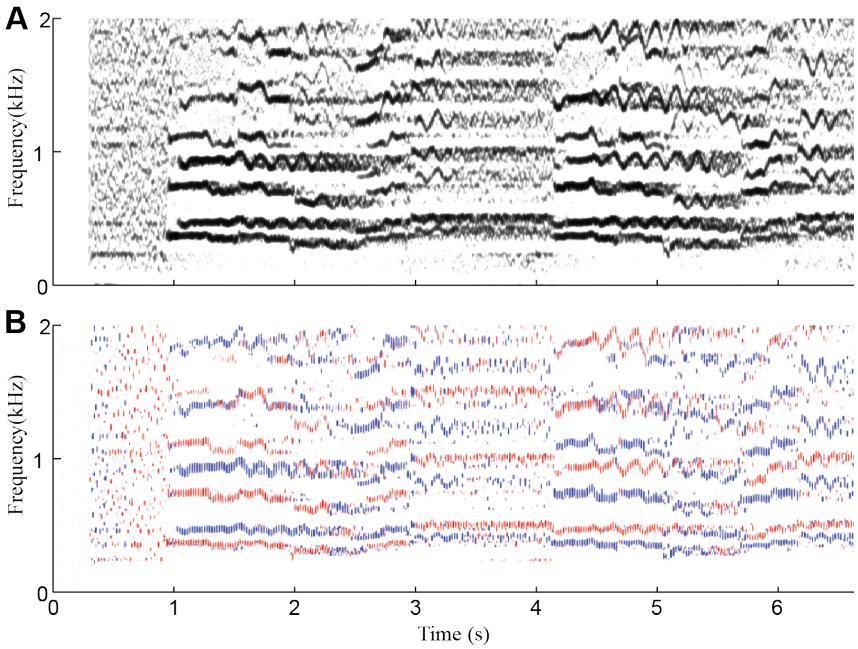

Figure 4.7: **Gamma partitioning of 'Flower Duet'. A.** Input stimulus composed of two vocal parts in the chorus. **B.** Gamma partitioning and streaming performed using gamma partitioning and k-means streaming.

4.4 Discussion

In this chapter I have considered how auditory partitioning might occur in a neural network using both recurrent excitatory and inhibitory connections. Partitioning occurs due to gamma wav oscillations, with different sets of perceptual objects represented in different gamma cycles. It has been shown how this partitioning works for a variety of auditory stimuli, including those which are relatively complex. Finally, I considered how this partitioning might be combined with a simple clustering algorithm to form continuous perceptual streams.

This partitioning of auditory stimuli is only possible because information about the stimulus is preserved in the same population of neurons. The recurrent excitatory

connections provide a feedback loop within each perceptual object, causing one object to be disproportionately represented in each cycle. The inhibitory recurrent connections aid in this, causing a 'winner-takes-all' environment which can only represent a small part of the stimulus. Inhibitory feedback is also crucial in resetting the cycle, so the network does not continuously represent one sole object. Finally, synaptic depression and neural refractory periods serve to penalize objects which have just been active, allowing the network to represent multiple successive objects over successive cycles. This also causes some phase-locking between neurons representing the same object, causing them to be more likely to fire simultaneously in future cycles. These processes all depend on the population of neurons maintaining a continued neural percept of the stimulus, continued over timescales longer than individual neural processing. They are therefore dependent on recurrent connectivity within the network.

The model presented shares several features with the LEGION network (Wang and Terman, 1995)(Section 1.7.4.2). The foremost amongst these is how segmentation is calculated using pre-learned features; this is well supported by psychoacoustic experiments on auditory streaming, which demonstrate how streaming is influenced by factors such as harmonicity, timbre and location. For example, the importance of harmonicity to auditory streaming can be gauged by considering whispered speech (which lacks harmonics); such speech can only be understood while it is relatively loud, and is quickly lost amongst background noise if the volume drops. These pre-learned features are represented in the recurrent connections, a premise broadly

supported by theories of long term plasticity (Section 1.3). Finally, like the earlier LEGION network, synchronicity between oscillators is used to distinguish different streams.

However, there are also several key differences between this network and earlier work. Some of these differences are caused by the different mechanisms underlying the network oscillations. In LEGION, oscillations are generated individually within each oscillating 'unit'. In contrast, oscillations are driven collectively by the global inhibitory neuron in the network presented above. Explicitly modeling all neural processes as standard LIF neurons does make the network presented above more biologically plausible. More importantly, however, global synchronization of oscillations means that synchrony is achieved much faster. In LEGION, all oscillators start out of phase and are slowly shifted in phase space until segments match, a process that takes at least 5-10 oscillations. In contrast, in the gamma wave network some level of synchrony is automatically imposed (global gamma waves), and clear partitioning is achieved within one or two oscillations. This allows the gamma wave network to segregate quickly evolving signals (Figure 4.6), a task that was not attempted with the earlier neural network streaming models.

The network presented in this chapter also expands on the concept of 'similar' neurons. Neurons can be considered to be similar if they are typically co-active in the presence of natural stimuli, or alternatively if there is a significant excitatory connection between them. Gamma partitioning essentially functions by clustering

together groups of 'similar' neurons. In the above network I consider two neurons to be 'similar' if they represent frequencies which might both occur in the same harmonic stack. Such a calculation is non-trivial; the definition of 'similar can no longer be thought to depend on a distance function (a metric) as the underlying function is not monotonic (200 Hz and 400 Hz are 'similar in harmonic space, whereas neither is 'similar' to 311 Hz). This level of complexity was not considered in the LEGION network, or in the paper which introduced the concept of gamma partitioning in the visual system (Miconi and Vanrullen, 2010). Indeed, when the LEGION model was later extended to consider harmonic stimuli this was achieved via extensive preprocessing to detect and gather frequencies belonging to harmonic stacks (Brown and Wang, 1997). In biological neural networks, each individual neuron may be sensitive to several different features. This means that two neurons might be fairly similar in some respects, but not in others. Consequently, the ability to cluster inputs using more complicated relationships between neurons is of great interest.

We also briefly considered how segments might be combined together into auditory streams. In particular, we applied a sliding k-means clustering algorithm. This algorithm was chosen as it is relatively simple, and has been linked to clustering in neural networks . This algorithm has trouble following streams when they undergo a sudden jump in features; this is not surprising given the temporal continuity basis. We expect stronger performance with the inclusion of more features, and with a larger neural representation; this model only considers 400 different receptive fields (albeit with 100 neurons with each set of inputs), far less than the 10^8 in the audi-

tory cortex. These features need not play an equal role in stream formation; some features may remain more stable within any given stream, and it is reasonable to assume that a neural network could weight features appropriately (and learn these weights through exposure to natural stimuli).

One interesting extension that could be explored is how this process of auditory streaming might be informed by my earlier work on transient attractors. It is highly plausible that the strengths of recurrent connections which dictate 'similarity' might depend on both 'hard-wired' correlations and also on correlations in the recent history. An attempt was made to incorporate transient attractors into the above auditory streaming model, although significant strengthening of connections was found to lead to runaway excitation within each gamma cycle. This in turn might be remedied through plastic inhibitory networks (Vogels and Abbott, 2009). This will hopefully prove to be an interesting area for future research.

While the gamma wave partitioning network presented in this chapter certainly falls short of a complete explanation of auditory streaming, this novel approach shows promise in its ability to perform complicated segmentation, in its biological realism, and in its potential for the incorporation of several complex features.

Methods

Simulations

All simulations were performed in MATLAB using the implicit Euler method with $\Delta t = 0.1$ ms.

Neuron Model

In all simulations I used a continuous firing rate model with cell voltage, $v_i(t)$, depending on weighted sum of recurrent excitatory, inhibitory and input currents,

$$\begin{aligned} C\frac{dV_j}{dt} &= I_j^{Leak}(t) + I_j^E(t) + I_j^I(t) + \alpha_N dW_t + In(t) \\ I_j^E(t) &= \sum_{i=1}^{N_E} g_{ij}^E(t) \\ I_j^I(t) &= \sum_{i=1}^{N_E} g_{ij}^I(t)[E_{rev}^I - V_j(t)] \\ I_j^{Leak}(t) &= g^{Leak}[V_{rest} - V_j(t)]. \end{aligned}$$

Here, dW_t is the time derivative of Brownian motion ($N(0, \sqrt{\Delta t})$). In this formulation, the excitatory input currents are independent of the cell membrane potential. This approximation is valid so long as the excitatory reversal potential is far higher than the firing threshold, which is the case for cortical neurons. In contrast, this can not be assumed for the inhibitory currents because the inhibitory reversal current is typically close to the cell membrane potential at rest.

Due to the spiking (dirac-delta) nature of the inputs, we model each of these con-

ductances as the product between a synaptic strength (a^*_{ij}) and a time course for the conductivity from each incoming spike,

$$g^*_{ij}(t) = W^*_{ij} \sum_{t_{spk}} \frac{1}{\tau^*_1 - \tau^*_2} \left[\exp\left(-\frac{t - t_{spk}}{\tau^*_1} \right) - \exp\left(-\frac{t - t_{spk}}{\tau^*_2} \right) \right] H(t - t_{spk})$$

(Gabbiani et al., 1994). The Heaviside step function (H) is included to ensure that the conductance may only change after each presynaptic spike occurs.

Parameters

Simulation parameters were constant across all simulations, and are listed in Table 2.1.

Other Packages

Spectrogram (DFT) were calculated using the 'myspectrogram.m' file by Kamil Wojcicki, Signal Processing Laboratory, Griffith University, Nathan, QLD, Australia, 2007. Accessed 2016. Available at https://www.mathworks.com/matlabcentral/fileexchange/29596-speech-spectrogram/content/myspectrogram/myspectrogram.m

Name	Symbol	Value
Maximum Synaptic Weights:		
Input to Exc	W_{SE}	0.02
Exc to Exc	W_{EE}	15
Exc to Local Inh	W_{EI}	0.5
Local Inh to Exc	W_{IE}	1
Exc to Global Inh	W_{EI_G}	0.2
Global Inh to Exc	W_{I_GE}	50
Synaptic timescales		
Excitatory synapse decr	τ_1^E	1 ms
Excitatory synapse incr	τ_2^E	0.2 ms
Excitatory synapse decr	τ_1^I	10 ms
Excitatory synapse incr	τ_2^E	2 ms
Other		
Leak current	g^{Leak}	0.1 ms^{-1}
Firing threshold	b	1
Resting current	V_{rest}	0
Inhibitory reversal potential	E_{rev}^I	-1
Noise	α_N	0.5

Table 4.1: **Gamma Partitioning Network parameters** Above parameters were used in all simulations in the chapter.

Chapter 5: Conclusions

In this dissertation, I have presented a set of neural network models which investigate the roles that recurrent connections in neural networks might play in processing sensory stimuli. These models all revolve around the idea of keeping an ongoing percept of recent stimuli within the sensory cortices, and using this to alter how the neural network processes any new incoming stimuli. This task is particularly important since the objects being processed in the sensory cortices typically persist for far longer than the timescales of neural processing, but far shorter than the timescales of long term synaptic modifications.

I started by investigating potential mechanisms behind short-term memory. This is one area in which recurrent neural connections have often been discussed; recurrent connections allow for stable attractors, and information may be stored by activating one such attractor. It is doubtful, however, that such a persistent activity model of short term memory explains all the capabilities of short-term memory. Instead, I investigated the role that temporary modifications to the recurrent network connectivity might play in storing memory. This plastic connectivity may not only emphasize pre-existing attractors, but also temporarily create attractors, which may

then be sensed using an appropriate prompt. I then went on to demonstrate how this concept of transient attractors could help in other tasks associated with sensory processing.

Next, I investigated the mechanisms which might allow for temporary modifications to the network's connectivity. Previous works have assumed some transient associative mechanism to store information, whereas none has been observed. Instead I investigated how this information might be stored by facilitating features within the network, making use of the process of synaptic facilitation and the hierarchical nature of stimulus processing. This alternative means of storing associative information might allow many different processes which depend on associative modifications to occur on a wider variety of timescales.

Finally, I considered the challenge of auditory segmentation, as the first step to auditory streaming. In line with synchronicity theory, I hypothesized that this segmentation might happen over multiple gamma cycles. This mechanism relies on two different recurrent mechanisms: the global inhibitory network creates a 'winner-takes-all' environment, and recurrent excitatory connections cause a clustering of neurons due to their pre-learned 'similarity'. This mechanism is also highly dependent on the continued representation of the stimulus within the same population of neurons; synaptic depression and refractory periods mean that a cluster of neurons which have just represented some auditory source are less likely to fire in the next gamma cycle.

Much work remains in understanding the role of recurrent connections in neural networks; for example, ideas from all of the above work might be formed into a more comprehensive model of auditory streaming. However, the work in my dissertation has demonstrated several new ways in which recurrent connections within the sensory cortices might aid in the processing of complex, natural stimuli.

Bibliography

Abbott, L. F. and Regehr, W. G. Synaptic computation. *Nature*, 431(7010):796–803, 2004.

Aksay, E., Baker, R., Seung, H. S., and Tank, D. W. Anatomy and discharge properties of pre-motor neurons in the goldfish medulla that have eye-position signals during fixations. *Journal of neurophysiology*, 84(2):1035–1049, 2000.

Aksay, E., Olasagasti, I., Mensh, B. D., Baker, R., Goldman, M. S., and Tank, D. W. Functional dissection of circuitry in a neural integrator. *Nature neuroscience*, 10 (4):494–504, 2007.

Amit, D. J., Fusi, S., and Yakovlev, V. Paradigmatic working memory (attractor) cell in IT cortex. *Neural computation*, 9(5):1071–92, 1997.

Amit, D. J., Gutfreund, H., and Sompolinsky, H. Storing infinite numbers of patterns in a spin-glass model of neural networks. *Physical Review Letters*, 55(14):1530–1533, 1985.

Arnal, L. H. and Giraud, A.-l. Cortical oscilations and sensory predictions. *Trends in cognitive sciences*, 16(7), 2012.

Atkinson, R. C. and Shiffrin, R. M. The Control Process of Short-Term Memory. Technical report, Institute for Mathematical Studies in the Social Sciences, 1971.

Attwell, D. and Laughlin, S. B. An energy budget for signaling in the grey matter of the brain. *Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism*, 21(10):1133–45, 2001.

Barak, O., Tsodyks, M., and Romo, R. Neuronal Population Coding of Parametric Working Memory. *Journal of Neuroscience*, 30(28):9424–9430, 2010.

Barak, O. and Tsodyks, M. Persistent activity in neural networks with dynamic synapses. *PLoS Computational Biology*, 3(2):0323–0332, 2007.

Barak, O. and Tsodyks, M. Working models of working memory. *Current Opinion in Neurobiology*, 25:20–24, 2014.

Becker, S. Learning to categorize objects using temporal coherence. Technical report, The Rotman Research Institute, 1992.

Becker, S. and Plumbley, M. Unsupervised Neural Network Learning Procedures For Feature Extraction and Classification. *Journal of Applied Intelligence*, 6, 1996.

Ben-Yishai, R., Bar-Or, R. L., and Sompolinsky, H. Theory of orientation tuning in visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 92(9):3844–3848, 1995.

Bendor, D. and Wang, X. The neuronal representation of pitch in primate auditory cortex. *Nature*, 436(7054):1161–1165, 2005.

Bizley, J. K. and Cohen, Y. E. The what, where and how of auditory-object perception. *Nature Reviews Neuroscience*, 14(10):693–707, 2013.

Bragin, A., Jandó, G., Nádasdy, Z., Hetke, J., Wise, K., and Buzsáki, G. Gamma (40-100 Hz) oscillation in the hippocampus of the behaving rat. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 15(1 Pt 1): 47–60, 1995.

Bregman, A. S. *Auditory Scene Analysis*. MIT Press, Cambridge, Mass, 1990.

Bregman, A. S. and Campbell, J. Primary Auditory Stream Segregation and Perception of Order in Rapid Sequences of Tones. *Journal of Experimental Psychology*, 89(2):244–249, 1971.

Brenowitz, S. D. and Regehr, W. G. Associative short-term synaptic plasticity mediated by endocannabinoids. *Neuron*, 45(3):419–431, 2005.

Brody, C. D. and Hopfield, J. J. Simple networks for spike-timing-based computation, with application to olfactory processing. *Neuron*, 37(5):843–52, mar 2003.

Brokx, J. P. L. and Nooteboom, S. G. Intonation and the perceptual separation of simultaneons voices. *Journal of Phonetics*, 10:23–36, 1982.

Brown, G. J. and Cooke, M. Computational auditory scene analysis, 1994.

Brown, G. J. and Wang, D. Modelling the perceptual segregation of double vowels with a network of neural oscillators. *Neural Networks*, 10(9):1547–1558, 1997.

Brunel, N. Dynamics and Plasticity of Stimulus-selective Persistent Activity in Cortical Network Models. *Cerebral Cortex*, 13(11):1151–1161, 2003.

Brunel, N. and Hakim, V. Fast Global Oscillations in Networks of Integrate-and-Fire Neurons with Low Firing Rates. *Neural Computation*, 11(7):1621–1671, 1999.

Buhmann, J. and Schulten, K. Associative recognition and storage in a model network of physiological neurons. *Biological Cybernetics*, 54(4-5):319–335, 1986.

Burkitt, A. N. A review of the integrate-and-fire neuron model: II. Inhomogeneous synaptic input and network properties. *Biological cybernetics*, 95(2):97–112, aug 2006a.

Burkitt, A. N. A review of the integrate-and-fire neuron model: I. Homogeneous synaptic input. *Biological cybernetics*, 95(1):1–19, jul 2006b.

Carlyon, R. P., Cusack, R., Foxton, J. M., and Robertson, I. H. Effects of attention and unilateral neglect on auditory stream segregation. *J. Exp. Psychol.: Human Percept. Perform.*, 27(1):115–127, 2001.

Casey, M. A. and Westner, A. Separation of mixed audio sources by independent subspace analysis. *Proceedings of the International Computer Music Conference*, pages 154–161, 2000.

Cherry, E. C. Some Experiments on the Recognition of Speech, with One and with Two Ears, 1953.

Courtney, S. M., Ungerleider, L. G., Keil, K., and Haxby, J. V. Transient and sustained activity in a distributed neural system for human working memory., 1997.

Curtis, C. E. and Lee, D. Beyond working memory: the role of persistent activity in decision making. *Trends in Cognitive Sciences*, 14(5):216–222, 2010.

Darwin, C. J. Auditory grouping. *Trends in cognitive sciences*, 1(9):327–333, 1997.

Darwin, C. J. and Bethell-Fox, C. E. Pitch continuity and speech source attribution. *Journal of Experimental Psychology: Human Perception and Performance*, 3(4): 665–672, 1977.

Dayan, P. and Abbott, L. F. *Theoretical Neuroscience*. MIT Press, Cambridge, Mass, 2001.

Delibes, L. *Flower Duet* from *Lakmé*. Performed by Anna Netrebko & Elina Garanca. [Video File]. Retrieved from https://www.youtube.com/watch?v=Vf42IP__ipw on 06/10/2016.

D'Esposito, M. and Postle, B. R. The Cognitive Neuroscience of Working Memory. *Annual Review of Psychology*, 66(1):115–142, 2015.

Deutsch, D. Binaural integration of melodic patterns. *Perception & psychophysics*, 25(5):399–405, 1979.

Ding, N., Chatterjee, M., and Simon, J. Z. Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. *NeuroImage*, 88: 41–46, 2014.

Doumont, J.-L. Magical Numbers : The Secen-plus-or-Minus-Two Myth. *Psychological Review*, 45(2):123–127, 2002.

Dranias, M. R., Ju, H., Rajaram, E., and VanDongen, A. M. J. Short-Term Memory in Networks of Dissociated Cortical Neurons. *Journal of Neuroscience*, 33(5): 1940–1953, 2013.

Druckmann, S. and Chklovskii, D. B. Over-complete representations on recurrent neural networks can support persistent percepts. *Advances in Neural Information Processing Systems*, 23:1–9, 2010.

Dubreuil, A. M. and Brunel, N. Storing structured sparse memories in a multi-modular cortical network model. *Journal of Computational Neuroscience*, 40(2): 157–175, 2016.

Dudai, Y. *Memory from A to Z*. Oxford University Press, Oxford, 2002.

Dudek, S. M. and Bear, M. F. Homosynaptic long-term depression in area CA1 of hippocampus and effects of N-methyl-D-aspartate receptor blockade. *Proceedings of the National Academy of Sciences of the United States of America*, 89(10): 4363–7, 1992.

Edin, F., Klingberg, T., Johansson, P., McNab, F., Tegnér, J., and Compte, A. Mechanism for top-down control of working memory capacity. *Proceedings of the National Academy of Sciences*, 106(16):6802–6807, 2009.

Ehret, G. The auditory cortex. *Journal of Comparative Physiology - A Sensory, Neural, and Behavioral Physiology*, 181(6):547–557, 1997.

Elhilali, M. and Shamma, S. a. A cocktail party with a cortical twist: how cortical mechanisms contribute to sound segregation. *The Journal of the Acoustical Society of America*, 124(6):3751–3771, 2008.

Engel, A. K., Koning, P., Gray, C. M., and Singer, W. Stimulus-dependent neuronal oscillations in cat visual cortex: Inter-columnar interaction as determined by cross-correlation analysis, 1990.

Engel, A. K. and Singer, W. Temporal binding and the neural correlates of sensory awareness. *Trends in Cognitive Sciences*, 5(1):16–25, 2001.

Erickson, M. A., Maramara, L. A., and Lisman, J. A single 2-spike burst induces GluR1-dependent associative short-term potentiation: a potential mechanism for short term memory. *Journal of cognitive neuroscience*, 22(11):2530–2540, 2011.

Erickson, R. *Sound Structure in Music*. University of California Press, Berkeley and Los Angeles, 1975.

Ermentrout, G. B. and Terman, D. H. *Mathematical Foundations of Neuroscience*, volume 35 of *Interdisciplinary Applied Mathematics*. Springer New York, New York, NY, 2010.

Felleman, D. J. and Van Essen, D. C. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex (New York, N.Y. : 1991)*, 1(1):1–47, 1991.

Fisher, S. A., Fischer, T. M., and Carew, T. J. Multiple overlapping processes underlying short-term synaptic enhancement. *Trends in Neurosciences*, 20(4): 170–177, 1997.

Fontolan, L., Krupa, M., Hyafil, A., and Gutkin, B. Analytical insights on theta-gamma coupled neural oscillators. *Journal of mathematical neuroscience*, 3:16, 2013.

Fourcaud, N. and Brunel, N. Dynamics of the Firing Probability of Noisy Integrate-and-Fire Neurons. *Neural computation*, 14:2057–2110, 2002.

Fries, P., Roelfsema, P. R., Engel, A. K., König, P., and Singer, W. Synchronization of oscillatory responses in visual cortex correlates with perception in interocular rivalry. *Proceedings of the National Academy of Sciences of the United States of America*, 94(23):12699–12704, 1997.

Fuster, J. M. and Alexander, G. E. Neuron Activity Related to Short-Term Memory, 1971.

Gabbiani, F., Midtgaard, J., and Knopfel, T. Synaptic Integration in a Model of Cerebellar Granule Cells. *Journal Of Neurophysiology*, 72(2):999–1009, 1994.

Ganguli, S., Huh, D., and Somploinsky, H. Memory traces in dynamical systems. *Proceedings of the National Academy of Sciences of the United States of America*, 105(48):18970–5, dec 2008.

Ghose, G. M. and Maunsell, J. Specialized representations in visual cortex: A role for binding? *Neuron*, 24(1):79–85, 1999.

Goldman, M. S. Memory without Feedback in a Neural Network. *Neuron*, 61(4): 621–634, 2009.

Goldman, M. S., Compte, A., and Wang, X.-j. Theoretical and computational neuroscience: Neural integrators: recurrent mechanisms and models. *Squire, L.; Albright, T.; Bloom, F*, 2007.

Gray, C. M., König, P., Engel, A. K., and Singer, W. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties., 1989.

Haykin, S. and Chen, Z. The Cocktail Party Problem. *Neural Computation*, 17(9): 1875–1902, 2005.

Hebb, D. O. *The Organization of Behavior*. Wiley, New York, 1949.

Hennig, M. H. Theoretical models of synaptic short term plasticity. *Frontiers in Computational Neuroscience*, 7(April):1–10, 2013.

Holcman, D. and Tsodyks, M. The emergence of up and down states in cortical networks. *PLoS Computational Biology*, 2(3):174–181, 2006.

Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, 79(April):2554–2558, 1982.

Hubel, D. H. and Wiesel, T. N. Receptive Fields, Binocular Interactions and Functional Architeture in the Cat's Visual Cortex. *J. Physiol*, 160:106–154, 1962.

Hubel, D. H., Wiesel, T. Receptive Fields of Single Neurons in the Cat's Striate Cortex. *Physiol, J*, 148:574–591, 1959.

Itskov, V., Hansel, D., and Tsodyks, M. Short-term facilitation may stabilize parametric working memory trace. *Frontiers in Computational Neuroscience*, 5 (October):1–19, 2011.

Izhikevich, E. M. *Dynamical Systems in Neuroscience: The Geometry of Excitability and Burtsing*. MIT Press, Cambridge, Mass., 2007.

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

Kaas, J. H., Hackett, T. A., and Tramo, M. J. Auditory processing in primate cerebral cortex. *Current Opinion in Neurobiology*, 9(2):164–170, 1999.

Kaiser, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23(3), 1958.

Kalman, R. E. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(1):35, 1960.

Kandel, E. R. Cellular Mechanisms of Learning and the Biological Basis of Individuality. In *Principles of Neural Science*, pages 1248–1280. McGraw-Hill Companies, Inc., 2014.

Keysers, C. and Perrett, D. I. Demystifying social cognition: A Hebbian perspective. *Trends in Cognitive Sciences*, 8(11):501–507, 2004.

King, P. D., Zylberberg, J., and DeWeese, M. R. Inhibitory Interneurons Decorrelate Excitatory Cells to Drive Sparse Code Formation in a Spiking Model of V1. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 33 (13):5475–5485, mar 2013.

Ko, H., Hofer, S. B., Pichler, B., Buchanan, K. A. K., Sjöström, P. J., and Mrsic-Flogel, T. D. Functional specificity of local synaptic connections in neocortical networks. *Nature*, 473(7345):87–91, 2011.

Krishnan, L., Elhilali, M., and Shamma, S. Segregating Complex Sound Sources through Temporal Coherence. *PLoS computational biology*, 10(12):1–10, dec 2014.

Kubota, K. and Niki, H. Prefrontal cortical unit activity and delayed alternation performance in monkeys. *Journal of neurophysiology*, 34(3):337–347, 1971.

Lansner, A., Fransen, E., and Sandberg, A. Cell Assembly Dynamics in Detailed and Abstract Attractor Models of Cortical Associative Memory. *Theory in Biosciences*, 122(1):19–36, 2003.

Larocque, J. J., Lewis-Peacock, J. a., and Postle, B. R. Multiple neural states of representation in short-term memory? It's a matter of attention. *Frontiers in human neuroscience*, 8(January):5, 2014.

Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., and Postle, B. R. Neural Evidence for a Distinction Between Short-Term Memory and the Focus of Attention. *Journal of cognitive neuroscience*, 24(1):61–79, 2012.

Lim, S. and Goldman, M. S. Balanced cortical microcircuitry for maintaining short-term memory. *Nature neuroscience*, 16(9):1306–1314, 2013.

Lisman, J. E. and Idiart, M. A. Storage of 7 +/- 2 short-term memories in oscillatory subcycles., 1995.

London Philharmonia. Sound Samples. `http://www.philharmonia.co.uk/explore/sound_samples`, 2016.

Lundqvist, M., Rose, J., Herman, P., Brincat, S. L., Buschman, T. J., and Miller, E. K. Gamma and Beta Bursts Underlie Working Memory. *Neuron*, 90(1):152–164, 2016.

Machens, C. K., Romo, R., and Brody, C. D. Flexible Control of Mutual Inhibition: A Neural Model of Two-Interval Discrimination. *Science*, 307(5712):1121–1124, 2005.

MacNeil, D. and Eliasmith, C. Fine-tuning and the stability of recurrent neural networks. *PLoS ONE*, 6(9), 2011.

Maex, R. and Steuber, V. The first second: Models of short-term memory traces in the brain. *Neural Networks*, 22:1105–1112, 2009.

Malenka, R. C. Postsynaptic factors control the duration of synaptic enhancement in area CA1 of the hippocampus. *Neuron*, 6(1):53–60, 1991.

Masquelier, T. Neural variability, or lack thereof. *Frontiers in computational neuroscience*, 7(February):7, 2013.

McDougal, R. A. *Excitatory-inhibitory interactions as the basis of working memory.* PhD thesis, Ohio State University, 2011.

Mejias, J. F. and Torres, J. J. Maximum memory capacity on neural networks with short-term synaptic depression and facilitation. *Neural computation*, 21(3): 851–71, 2009.

Mercury, F. Bohemian Rhapsody. Performed by Queen, 1975. [Video File]. Retrieved from https://www.youtube.com/watch?v=lXZhmvGusfs on 06/10/2016.

Miconi, T. and Vanrullen, R. The gamma slideshow: object-based perceptual cycles in a model of the visual cortex. *Frontiers in human neuroscience*, 4:205, jan 2010.

Miller, G. A. The Magical Number Number Seven, Plus of Minus Two: Some Limits on Our Capacity for Processing Information. *The Psychological Review*, 63:81–97, 1956.

Miller, G. A. MIller, George A. In Lindzey, G., editor, *A History of Psychology in Autobiography*, page 401. Stanford University Press, 1989.

Miller, G. A. and Heise, G. A. The Trill Threshold. *Journal of the Acoustical Society of America*, 22(5):637–638, 1950.

Mongillo, G., Barak, O., and Tsodyks, M. Synaptic theory of working memory. *Science (New York, N.Y.)*, 319:1543–1546, 2008.

Movshon, J. A. Reliability of neuronal responses. *Neuron*, 27(3):412–414, 2000.

Murtagh, F. and Hernández-Pajares, M. The Kohonen self-organizing map method: An assessment. *Journal of Classification*, 12(2):165–190, 1995.

Olivers, C. N. L., Peters, J., Houtkamp, R., and Roelfsema, P. R. Different states in visual working memory: When it guides attention and when it does not. *Trends in Cognitive Sciences*, 15(7):327–334, 2011.

Olshausen, B. A. and Field, D. J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381, 1996.

Olshausen, B. and Field, D. What is the other 85 % of V1 doing? *Problems in Systems . . .*, pages 1–29, 2004.

Orhan, E. The Hopfield Model. Technical report, NYU, 2014.

Oswald, J. P., Klug, A., and Park, T. J. Interaural intensity difference processing in auditory midbrain neurons: effects of a transient early inhibitory input. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 19 (3):1149–1163, 1999.

Pasternak, T. and Greenlee, M. W. Working memory in primate sensory systems. *Nature reviews. Neuroscience*, 6(2):97–107, 2005.

Peters, A., Payne, B. R., and Budd, J. A numerical analysis of the geniculocortical input to striate cortex in the monkey. *Cerebral Cortex*, 4(3):215–229, 1994.

Petrides, M. Dissociable roles of mid-dorsolateral prefrontal and anterior inferotemporal cortex in visual working memory. *J Neurosci*, 20(19):7496–7503, 2000.

Postle, B. R. Neural Bases of the Short-Term Retention of Visual Information. In *Attention & Performance XXV: Mechanisms of Sensory Working Memory*, pages 43–58. Elsevier, 2015.

Purves, D., Augustine, G. J., Fitzpatrick, D., Hall, W. C., LaMantia, A.-S., and White, L. E. *Neuroscience*. Sinauer Associates, Inc., Suderland, MA, fifth edition, 2012.

Qin, L., Chimoto, S., Sakai, M., Wang, J., and Sato, Y. Comparison Between Offset and Onset Responses of Primary Auditory Cortex. *Journal of Neurophysiology*, pages 3421–3431, 2007.

Qin, M. K. and Oxenham, A. J. Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. *The Journal of the Acoustical Society of America*, 114(1):446–454, 2003.

Regehr, W. G. Short-term presynaptic plasticity. *Cold Spring Harbor perspectives in biology*, 4(7):a005702, 2012.

Riesenhuber, M. and Poggio, T. Hierarchical models of object recognition in cortex. *Nature neuroscience*, 2(11):1019–25, 1999.

Roelfsema, P. R., Engel, A. K., König, P., and Singer, W. Visuomotor integration is associated with zero time-lag synchronization among cortical areas., 1997.

Rolls, E. T., Dempere-Marco, L., and Deco, G. Holding Multiple Items in Short Term Memory: A Neural Mechanism. *PLoS ONE*, 8(4), 2013.

Rose, J. E., Gross, N. B., Geisler, C. D., and Hind, J. E. Some neural mechanisms in the inferior colliculus of the cat which may be relevant to localization of a sound source. *Journal of neurophysiology*, 29(2):288–314, 1966.

Roskies, A. L. The binding problem. *Neuron*, 24(1):7–9, 1999.

Rozell, C. J., Johnson, D. H., Baraniuk, R. G., and Olshausen, B. A. Sparse coding via thresholding and local competition in neural circuits. *Neural computation*, 20 (10):2526–63, oct 2008.

Sandamirskaya, Y., Zibner, S. K. U., Schneegans, S., and Schöner, G. Using Dynamic Field Theory to extend the embodiment stance toward higher cognition. *New Ideas in Psychology*, 31(3):322–339, 2013.

Sandberg, A., Tegnér, J., and Lansner, A. A working memory model based on fast Hebbian learning. *Network*, 14(4):789–802, 2003.

Schneegans, S. and Schöner, G. *Dynamic field theory as a framework for understanding embodied cognition*. Elsevier Inc., 2008.

Schnupp, J., Nelken, I., and King, A. *Auditory Neuroscience*. MIT Press, Cambridge, Mass, 2011.

Schultz, W. and Dickinson, A. Neuronal Coding of Prediction Errors. *Annual review of neuroscience*, 23:473–500, 2000.

Serre, T., Wolf, L., and Poggio, T. Object Recognition with Features Inspired by Visual Cortex. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2:994–1000, 2005.

Seung, H. S. How the brain keeps the eyes still. *Proceedings of the National Academy of Sciences of the United States of America*, 93(23):13339–13344, 1996.

Shafi, M., Zhou, Y., Quintana, J., Chow, C., Fuster, J., and Bodner, M. Variability in neuronal activity in primate cortex during working memory tasks. *Neuroscience*, 146(3):1082–1108, 2007.

Shamma, S. A., Elhilali, M., and Micheyl, C. Temporal coherence and attention in auditory scene analysis. *Trends in Neurosciences*, 34(3):114–123, 2011.

Shamma, S. A., Elhilali, M., Ma, L., Micheyl, C., Oxenham, A. J., Pressnitzer, D., Yin, P., and Xu, Y. Temporal Coherence and the Streaming of Complex Sounds. *Adv Exp Med Biol.*, 787:535–543, 2013.

Shao, Y. and Wang, D. Sequential organization of speech in computational auditory scene analysis. *Speech Communication*, 51(8):657–667, 2009.

Shatz, C. The developing brain. *Scientific American*, 1992.

Shen, L. Neural Integration by Short Term Potentiation. *Biological Cybernetics*, 61: 319–325, 1989.

Singer, W. and Gray, C. M. Visual feature integration and the temporal correlation hypothesis. *Annual review of neuroscience*, 18:555–586, 1995.

Smaragdis, P. Non-negative matrix factor deconvolution; extraction of multiple sound sources from monophonic inputs. Technical Report 2, Mistubishi Electric Research Laboratories, 2004.

Song, S., Sjostrom, P. J., Reigl, M., Nelson, S., and Chklovskii, D. B. Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biology*, 3(3):0507–0519, 2005.

Sreenivasan, K. K., Curtis, C. E., and D'Esposito, M. Revisiting the role of persistent neural activity during working memory. *Trends in Cognitive Sciences*, 18(2): 82–89, 2014.

Stokes, M. G. Activity-silent' working memory in prefrontal cortex: a dynamic coding framework. *Trends in Cognitive Sciences*, 19(7):394–405, 2015.

Stokes, M. G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., and Duncan, J. Dynamic coding for cognitive control in prefrontal cortex. *Neuron*, 78(2):364–375, 2013.

Sugase-Miyamoto, Y., Liu, Z., Wiener, M. C., Optican, L. M., and Richmond, B. J. Short-term memory trace in rapidly adapting synapses of inferior temporal cortex. *PLoS computational biology*, 4(5):e1000073, 2008.

Sussman, E. Integration and segregation in auditory scee analysis. *J Acoust Soc Am*, 117(3), 2005.

Szatmáry, B. and Izhikevich, E. M. Spike-Timing Theory of Working Memory. *PLoS Computational Biology*, 6(8):e1000879, 2010.

Teki, S., Chait, M., Kumar, S., Shamma, S., and Griffiths, T. D. Segregation of complex acoustic scenes based on temporal coherence. *eLife*, 2013(2):1–16, 2013.

Terman, D. H. and Wang, D. Global competition and local cooperation in a network of neural oscillators. *Physica D*, 81:148–176, 1995.

Tetzlaff, C., Kolodziejski, C., Markelic, I., and Wörgötter, F. Time scales of memory, learning, and plasticity. *Biological Cybernetics*, 106(11-12):715–726, 2012.

Theunissen, F. E. and Elie, J. E. Neural processing of natural sounds. *Nature reviews. Neuroscience*, 15(6):355–66, 2014.

Tsodyks, M., Pawelzik, K., and Markram, H. Neural Networks with Dynamic Synapses. *Neural Computation*, 10, 1998.

Varela, F. J., Thompson, E., and Rosch, E. *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press, 1991.

Venkatasubramanian, S. Clustering on Streams. In *Encyclopedia of Database Systems*. Springer, 2009.

Verduzco-Flores, S., Bodner, M., Ermentrout, G. B., Fuster, J. M., and Zhou, Y. Working memory cells' behavior may be explained by cross-regional networks with synaptic facilitation. *PloS one*, 4(8):e6399, 2009.

Vogels, T. P. and Abbott, L. F. Gating multiple signals through detailed balance of excitation and inhibition in spiking networks. *Nature neuroscience*, 12(4):483–91, apr 2009.

Von der Malsburg, C. and Schneider, W. A neural cocktail-party processor. *Biological Cybernetics*, 54(1):29–40, 1986.

Von der Malsburg, C. The correlation theory of brain function. Technical report, Max-Planck Institute, 1981.

Wang, D. Primitive auditory segregation based on oscillatory correlation. *Cognitive Science*, 20(3):409–456, 1996.

Wang, D. and Chang, P. An oscillatory correlation model of auditory streaming. *Cognitive Neurodynamics*, 2(1):7–19, 2008.

Wang, D. and Terman, D. H. Local Excitatory Global Inhibitory Oscillator Networks. *IEEE Transactions on Neural Networks*, 6(1), 1995.

Wang, D. L. and Brown, G. J. Separation of speech from interfering sounds based on oscillatory correlation. *IEEE Transactions on Neural Networks*, 10(3):684–697, 1999.

Wang, X.-j., Introduction, I., Synchronization, A., Resonance, B., Subthreshold, C., Rhythms, V. S., Irregular, W., Activity, N., Communication, L.-d., and Learning, C. Neurophysiological and Computational Principles of Cortical Rhythms in Cognition. *Physiol Rev*, 90:1195–1268, 2010.

Watanabe, K. and Funahashi, S. Prefrontal delay-period activity reflects the decision process of a saccade direction during a free-choice ODR task. *Cerebral Cortex*, 17:i88–i100, 2007.

Watt, A. J. and Desai, N. S. Homeostatic Plasticity and STDP: Keeping a Neuron's Cool in a Fluctuating World. *Frontiers in synaptic neuroscience*, 2(June):5, jan 2010.

Wei, Z., Wang, X.-J., and Wang, D.-H. From Distributed Resources to Limited Slots in Multiple-Item Working Memory: A Spiking Network Model with Normalization. *The Journal of Neuroscience*, 32(33):11228–11240, 2012.

Weinberger, N. M. Plasticity in the Primary Auditory Cortex, Not What You Think it is: Implications for Basic and Clinical Auditory Neuroscience. *Otolaryngol (Sunnyvale)*, 4(164):1–19, 2012.

Wessel, D. Timbre Space as a Musical Control Structure. *Computer Music Journal*, 3(2):45–52, 1979.

Wimmer, K., Nykamp, D. Q., Constantinidis, C., and Compte, A. Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nature neuroscience*, 17(3):431–9, 2014.

Worden, F. G. Hearing and the Neural Detection of Acoustic Patterns. *Behavioral Science*, 16(1), 1971.

Wrigley, S. N. and Brown, G. J. A Computational Model of Auditory Selective Attention. *IEEE Transactions on Neural Networks*, 15(5):1151–1163, 2004.

Wurtz, R. H. Recounting the impact of Hubel and Wiesel. *The Journal of physiology*, 587(Pt 12):2817–2823, 2009.

Zarahn, E., Aguirre, G., and Esposito, M. D. A Trial-Based Experimental Design for fMRI. *Neuroimage*, 6:122–138, 1997.

Zucker, R. S. and Regehr, W. G. Short-Term Synaptic Plasticity. *Annual Review of Physiology*, 64(1):355–405, 2002.

Zylberberg, J., Murphy, J. T., and DeWeese, M. R. A sparse coding model with synaptically local plasticity and spiking neurons can account for the diverse shapes of V1 simple cell receptive fields. *PLoS computational biology*, 7(10):e1002250, oct 2011.