Recurrent network dynamics in visual cortex:

a neural mechanism for spatiotemporal integration

By Jeroen Joukes

A dissertation submitted to the

Graduate School-Newark Rutgers, The State University of New Jersey

in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Graduate Program in Behavioral and Neural Sciences

written under the direction of Bart Krekelberg

and approved by:

Dr. Farzan Nadim ___________________

Dr. Horacio Rotstein ___________________

Dr. Tibor Koos ___________________

Dr. Jonathan D. Victor ___________________

Dr. Bart Krekelberg ___________________

Newark, New Jersey

October, 2016

ABSTRACT OF THE DISSERTATION

Recurrent network dynamics in visual cortex:

a neural mechanism for spatiotemporal integration

By: Jeroen Joukes

Thesis director: Bart Krekelberg

We present a data-driven computational approach for studying neural systems. In this approach one starts with experimental stimuli (inputs) and measured neuronal responses (outputs). The relationship between the inputs and outputs is modeled with an artificial recurrent neural network (ARNN). A detailed investigation of the network weights and response properties of the connected elements, together with simulated experiments performed on the ARNN leads to significant new insights and new hypotheses about the underlying neural mechanisms.

We first applied this approach to motion responses of neurons in the macaque middle temporal area (MT). This provided the novel insight that recurrent networks dynamics can explain complex motion tuned response dynamics found in MT neurons, without the need for feedforward temporal delay lines.

In our second study we used this approach to model the early visual form processing pathway of the macaque brain. Neurons in the secondary visual cortex (V2) were stimulated with textured stimuli designed to probe the visual systems for complex visual shapes. The approach led to the novel hypothesis that selectivity for complex form depends on selectivity for motion.

For the third study we extended the approach by taking advantage of chronically implanted microelectrode arrays (FMA) in primary visual cortex (V1) of the awake behaving macaque. With the FMA we collected V1 responses on day one, fitted an ARNN, explored the detailed properties of the ARNN the following days, and tested model predictions with a V1 validation experiment within the same week. We found that V1 selectivity for form is much more complex than commonly thought and includes spatiotemporal interactions between multiple hotspots in the receptive field. With this approach we found complex V1 tuning properties that are currently thought to primarily arise higher up in the visual processing stream.

We conclude that ARNNs can offer a useful tool set for systems neuroscience; the powerful computational approach, together with carefully designed experiments, provides novel hypotheses and insights into the complexity of neural function.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

1 GENERAL INTRODUCTION

Computational modeling is an invaluable tool to analyze complex dynamical systems; it allows us to visualize, abstract, and systematically conceptualize large and often noisy data sets that may otherwise be beyond our comprehension. Fueled by the exponential increase of raw computational power, methods have been designed and fine-tuned to aid in our understanding of the billions of connected neurons that make up the human brain. Computational models in the field of systems neuroscience are dominated by what we will call a 'rule-based' approach. The computational neuroscientist builds a model from the ground up, inspired by current knowledge about the neural system of interest that is often phrased qualitatively in experimental papers ("area X is connected to area Y", "neuron X responds to feature Y in the environment"). These are the rules that the modeler converts into the quantitative properties of the model. Such a model can become a proof of principle showing that the set of rules is sufficient to generate a certain input/output behavior, and it can be used to investigate which of the rules are necessary (by modifying components of the model).

The main advantage of the rule-based approach is that the designer has a relatively complete understanding of the basic properties that underlie the model. As a consequence, predictions of the model can often be traced back to one or more of those basic properties. Of course, this is only possible if the number of free parameters (or the number of rules) is small. One significant disadvantage of this approach, however, is that only regularities that can easily be phrased as ‘rules’ are implemented in a model. As a consequence, such models tend to oversimplify the complexity of neural processing. Moreover, because the models are built on an interpretation of experimental data (not necessarily the data themselves), they tend to confirm interpretations, rather than challenge them.

A typical rule-based model we will discuss in detail in section 1.1.1 is the motion energy (ME) model (Adelson & Bergen, 1985; Watson & Ahumada, 1985). The ME model consists of a feedforward chain of neural operations that collectively compute the speed and direction of motion of a moving object. Each stage in the ME model is well defined (Figure 2) and the model generates predictions that can be tested experimentally with behavioral and electrophysiological methods. Several research groups have confirmed key predictions made by the ME model and this has led to a widely accepted view that motion processing in humans roughly follows the feedforward chain of operations proposed by the ME model. However, to compute velocity, any motion detector must compare visual input at two instances of time. In the ME model this is achieved by feedforward delay lines; the visual input to one set of units is temporally delayed compared to the input to another set of units. Even though the primate visual system does contain classes of slow and faster responding neurons (the parvo- and magnocellular stream, respectively), the evidence that this arrangement underlies the temporal delays required in a motion detector is controversial (Colby, 1981; De Valois & Cottaris, 1998; Maunsell & Nealey, 1990; De Valois, Cottaris, Mahon, Elfar, & Wilson, 2000) and the delay line is a feature of the ME model that is assumed to exist but not explained.

Conversely, recurrent connections are ubiquitous within and between all cortical brain areas, but this feature is ignored by the strictly feedforward ME model. One might expect that such fundamental differences between a model and a neural system should be reflected in discrepancies between model and neural response properties. How then is it possible that feedforward models, such as the ME model, nevertheless perform so well? The core underlying reason is that feedforward models can approximate any input-output mapping arbitrarily well (Funahashi & Nakamura, 1993; Hornik, Stinchcombe, & White, 1988). More specifically, consider the reverse correlation method (Marmarelis & Marmarelis, 1978). With this powerful technique one can approximate any input-output relationship using a feedforward model. It has been used to estimate feedforward models for neurons of the motion processing pathway and the properties of these models showed remarkable similarities with various predictions made by the ME model (Conway, 2003; Livingstone, Pack, & Born, 2001; Mechler & Ringach, 2002; Rust, Schwartz, Movshon, & Simoncelli, 2005; De Valois et al., 2000). This certainly suggests that elements of the ME model play a vital role in motion detection in the human brain. However, it is incorrect to deduce from this that the underlying neural mechanisms of motion detection are feedforward, because any architecture, any truly underlying mechanism, could be approximated with a feedforward method.

In this thesis, we present a fundamentally different data-driven approach for studying neural systems. In this approach one starts with actual data, not the "rule" that the experimentalist has used to describe or summarize the data. The relationship between experimental inputs and neural outputs (i.e. the data) is modeled using an artificial neural network (ANN, Figure 3). The ANN consists of many units that each can be considered a crude approximation of a population of neurons (cell assemblies). The units are interconnected with modifiable weights that loosely correspond to the effective synaptic connectivity between populations of neurons in the brain. During the fitting process (training phase) the weights are adjusted until the ANN reproduces a desired input-output relationship. It is important to realize that many ANNs are so-called universal approximators; they can in principle fit any input onto any output. This implies that capturing the input-output relationship in the experimental data on its own does not provide any meaningful insight, nor does it provide support for the model. As we will show below, significant insight can, however, be gained after the training phase with a detailed investigation of the ANN. In other words, in the data-driven approach, the hardware of the model is only very loosely constrained by specific experimental data or our general understanding of neural systems. Experimental data, however, are used to adjust the software (the weights) until a solution is found to map the input to the output. Insight is gained primarily from investigating this working solution. Because the model is constructed without imposing preconceived ideas about how the input should be mapped to the output, it is less likely to confirm preconceived ideas, and more likely to result in novel hypotheses and experimental predictions.

In chapter 2 we show the results of a case study of our ANN approach that we applied to motion tuned neurons recorded in the middle temporal area (MT) of the macaque brain (Joukes, Hartmann, & Krekelberg, 2014). In chapter 3 we build upon the case study and apply our ANN approach to neurons that were recorded in secondary visual cortex (V2) when stimulated with textured stimuli designed to probe the visual systems for complex visual features. Both models provided novel insights into how motion and form detection were solved by the models. However, the experimental data was collected with single cell recordings and this did not allow for experiments with the same neurons to test the model predictions. In chapter 4 we eliminated this limitation by taking full advantage of chronically implanted multiunit floating microelectrode arrays (FMA) in primary visual cortex (V1) of the awake behaving macaque that can record the same set of neurons over multiple recordings sessions spanning several weeks or even months. This allowed us to generate surprising predictions of fine-scale spatio-temporal interactions, and confirm them with follow-up experiments.

First, we will provide a brief introduction of the neural processing within the brain areas we modeled with the three projects. Given that ANNs play such an important role in this approach, we will also cover the basics of ANNs and describe how such networks can be used to model neural data.

1.1 MOTION PROCESSING

Anticipating moving predators and prey has a clear survival advantage and this evolutionary pressure has led to neural mechanisms of motion detection in many species throughout the animal kingdom. Much of our current knowledge about the neural mechanism of motion detection is based on behavioral and electrophysiological experiments with nonhuman primates whose brain areas involved in motion processing are similar to those in humans (Orban et al., 2003).

When a moving stimulus is presented to the eye it activates photoreceptors in the retina. Retinal ganglion cells pass the signal through the optic nerve and optic tract to the lateral geniculate nucleus of the thalamus (LGN). The LGN projects heavily to the primary visual cortex (V1), the largest cortical area of the macaque brain (Felleman & Van Essen, 1991). V1 neurons have tuning to low level visual features such as mean luminance, spatial scale, orientation and color. In addition, a substantial proportion of the neurons are tuned to a specific speed and direction of the moving stimulus. For a more detailed review of V1 we refer to (Callaway, 1998; Merigan & Maunsell, 1993; Sincich & Horton, 2005). V1 has strong projections to area MT, the motion processing area of the macaque brain (Maunsell & van Essen, 1983). This extrastriate cortical area is homologous to V5 in humans and here almost all neurons are velocity tuned. The speed tuned and direction selective response properties of MT neurons were discovered in the early 70’s by (Dubner & Zeki, 1971; Kaas, 1971) and many studies have confirmed the existence of a motion selective area in several species of both New and Old World monkeys (Baker, Petersen, Newsome, & Allman, 1981; Felleman & Kaas, 1984; Zeki, 1974). Finally, area MT is connected to many other subcortical and cortical brain areas; it receives direct input from the LGN (Sincich, Park, Wohlgemuth, & Horton, 2004), the inferior pulvinar (Warner, Goldshmit, & Bourne,

2010) the secondary visual cortex (V2), dorsal V3 (Felleman & Van Essen, 1991; Ungerleider & Desimone, 1986) and the medial superior temporal (MST). In addition, area MT is, similar to almost all cortical areas, dominated by recurrent network connectivity (Maunsell & van Essen, 1983). For a more detailed review of the various brain areas and network connectivity involved in motion processing, we refer to (Livingstone et al., 2001; Sincich & Horton, 2005).

Figure 1 shows the response of a typical MT neuron during the presentation of random dot patterns. The patterns were centered on the receptive field (RF) of the MT neuron and moved with one of seven speeds in the preferred and anti-preferred motion direction for a duration of 500 ms (Joukes et al., 2014). The mean response over time shows that the MT cell responded more for almost all tested speeds in the preferred motion direction (i.e. the neuron was direction selective) with a peak response for 16 °/s (its preferred speed). Figure Figure 1B shows the temporal dynamics of the response averaged over the seven speeds in the preferred direction (black line) and in the anti-preferred direction (gray line) for the first 250 ms after stimulus onset. After an onset delay of 27 ms direction selectivity started to emerge, it reached a relatively stable peak 93 ms after stimulus onset.

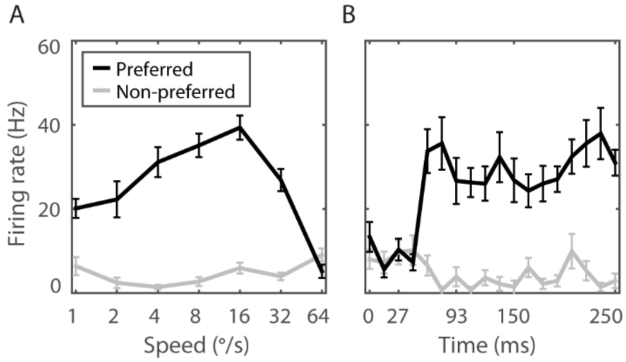

Figure 1. Motion tuned MT neuron. A) Mean response over time of an example MT neuron to random dot patterns moving at one of seven speeds (horizontal axis) in the preferred (black line) and anti-preferred direction (gray line). Error bars indicate standard error over trials (spatial patterns) per motion condition. B) MT cell response over time (x-axis) shown as the response averaged over seven speeds in the preferred direction (black line) and anti-preferred direction (gray line). Error bars indicate standard error over trials.

1.1.1 Motion detection and the motion energy model

In its simplest form, motion is defined as the spatial displacement of an object over time. A neural mechanism for motion detection, therefore, must have a mechanism for object localization (x), keeping track of spatial displacements (dx) and registering the duration of the spatial displacements (dt). A motion detector combines these building blocks such that the velocity (dx/dt) can be estimated. To illustrate this, we start with a more convenient visual representation of motion; a space-time orientation map (Figure 2, bottom space-time graph). Here, motion is defined as a slant in space-time where the slope and direction of the slant are the speed and direction of motion of the moving object (Figure 2, thick black line). Motion detection can now be defined as a mechanism that is capable of detecting the slope and direction of this space-time slant.

Figure 2 shows a schematic overview of the most dominant view of motion processing, the motion energy (ME) model (Adelson & Bergen, 1985; Watson & Ahumada, 1985) where motion signals are processed in a strictly feedforward manner. For a more complete description we refer to the original work by (Adelson & Bergen, 1985) or the review by (Krekelberg, 2008). The ME model consists of a feedforward chain of neural operations that compute motion energy (Figure 2, from bottom to top). The input units of the ME model (Figure 2, the red and blue colored circles) detect light reflected from the moving object at different spatial locations (dx) and they respond with different temporal delays (dt between solid and dashed lines in Figure 2). A summation over the outputs of the two red colored light sensors that are perfectly aligned with the space-time slant will, therefore, respond maximally to the speed and direction of motion of the moving object.

By changing dx and/or dt detectors can be made sensitive to different speeds and directions. The blue colored light sensors in Figure 2, for instance, respond most to leftward motion. The next stage of the ME model sums the output of the two similarly colored detectors to generate a unit that prefers a specific slanted space-time pattern (Figure 2). This space-time filter or response map can be read as a stimulus sensitivity profile; space-time input stimuli that perfectly match the response map elicit high activation and space-time input stimuli with a less optimal slope or slant elicit low activation at this ME model stage. In other words, the unit is sensitive to the speed and direction of motion defined by the dx/dt of the slanted space-time filter.

The space-time filter also shows that the output of the two light sensors cannot account for two essential properties of motion perception; phase and contrast invariance. First, phase invariance refers to the phenomenon that motion should not be dependent on where in the receptive field an object moves. Phase invariance is solved by the ME model with an additional set of spatial detectors with identical dx/dt properties of the red (or blue) colored set but spatially shifted to the left or to the right (not shown in Figure 2). A simple summation over the combined output of the two sets assures phase invariance. Second, contrast invariance refers to the phenomenon that the percept of motion does not depend on whether the moving object is white on a black background or black on a white background. Contrast polarity invariance is solved in the ME model by squaring the combined outputs of the two detector sets, assuring identical outputs for motion inputs of both contrast polarities. Finally, the blue colored detectors in Figure 2 were introduced in the ME model to increase direction selectivity. The space-time properties of the blue colored detectors are such that they respond maximally for motion in the opposite direction of the red colored detectors. Therefore, subtracting the response of the red and blue colored detector sets in the opponency stage, amplifies direction selectivity of the motion detector.

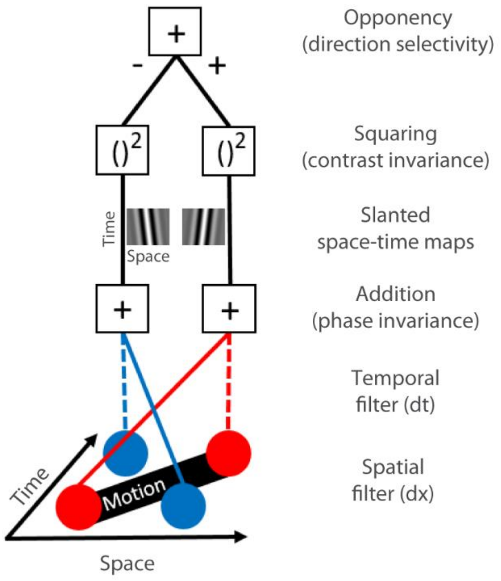

Figure 2. The Motion Energy (ME) model. A) ME model. The slanted space-time map (bottom) defines the speed and direction of a moving object (thick black bar). Shifted light sensors (red and blue circle sets) filter the motion signal in space with dx and dt. The spatial filter is typically implemented with Gabors with spatially shifted peaks. The temporal filter in the ME model is implemented with feedforward temporal delay lines (dashed lines). Light detectors with temporal delay lines transfer the motion signal slower than the lights detectors with the solid lines. The red detector set is oriented in space-time such that the sum over the spatial and temporal filter is maximal for the speed and direction defined by the slant in the space-time response maps (dx/dt). The output of each detector set is then squared to generate the same output for black objects moving on a white background and white objects moving on a black background (contrast invariance). Finally, in the opponency stage, direction selectivity is increased with a subtraction of the detectors that are sensitive to the opposite direction of motion. The ME model can generate motion tuning curves similar to those shown in Figure 1.

In this section we gave an overview of the current state of the field where motion processing is dominated by (various versions of) the ME model. Even though the model captures many neural properties, we will show in chapter 2 how our ANN approach generates new insights that challenge a strictly feedforward view of motion processing.

1.2 FORM PROCESSING

More than half a century ago (Hubel & Wiesel, 1962) discovered that neurons in cat primary V1 have selectivity for specific orientation of visual stimuli; they were orientation tuned. Some cells were tuned to either a white or a black oriented bar, but not both; they called them simple cells. Hubel and Wiesel also discovered neurons that were selective for oriented bars of both polarities (the same response for white and black oriented bars); they called them complex cells. Based on their pioneering experimental findings they proposed a model for simple and complex cells. They hypothesized that the tuning of simple cells could be constructed by summing the outputs of several lateral geniculate nucleus (LGN) cells. The LGN is part of the visual processing stream, has strong projections to V1 and contains cells that have small circular center-surround RFs. A simple summation over the outputs of a few neighboring LGN cells with RFs along a line can build orientation selectivity for a V1 simple cell. The tuning properties of complex cells are, in turn, constructed with a summation over at least two V1 simple cells with overlapping RFs, tuned to the same orientation, but sensitive to opposite contrasts. This feedforward view of orientation tuning and strict dichotomy of simple and complex cells has dominated the field of early visual form processing.

More complex visual scenes are processed hierarchically over many visual cortical areas in the brain. Whereas V1 simple and complex cells are thought to primarily respond to basic visual features such as mean luminance and oriented gratings, neurons further downstream the form processing pathway are sensitive to more complex shapes. The secondary visual cortex (V2) is the largest extrastriate visual cortical area in the macaque brain that receives its main input from V1 (Sincich & Horton, 2005). It is, therefore, not surprising that V2 shares many properties with V1 such as a sensitivity for luminance and oriented gratings (Heydt & Peterhans, 1989; De Valois et al., 2000). The response properties to more complex images revealed that V2 neurons had sensitivities for (illusory) contours (Heydt & Peterhans, 1989; De Valois et al., 2000), combinations of orientations (Anzai, Peng, & Van Essen, 2007) and angles, arcs and circles (Hegdé & Van Essen, 2000).

In the motion domain, parameters such as speed and motion direction can be adjusted and, therefore, investigated systematically. In form processing such parametrization is less intuitive, maybe in part because the concept itself is much more flexible and ill defined. For instance, researchers have resorted to documenting the response to a wide range of different shapes, but the choice of the shapes has been somewhat arbitrary or ad-hoc (Hegdé & Van Essen, 2000). Furthermore, investigating specific sensitivities of neurons along the form processing pathway is complicated by sensitivities of the same neurons to more simplistic visual features. For instance, a neuron that can detect contours is often also sensitive to oriented lines or mean luminance.

One solution for a more formal parametrized approach was applied to macaque single cell recordings and human functional magnetic resonance imaging (Freeman, Ziemba, Heeger, Simoncelli, & Movshon, 2013). In this study, artificial image textures were synthesized from natural photographs. These images had complex visual features but were designed to share mean luminance and spatial frequency content. With these image sets the visual system can be selectively targeted for complex tuning properties without the confound of a sensitivity for lower order features. With this study, a distinct functional role for V2 was discovered; whereas V1 neurons responded only marginally to the complex images, V2 showed a much more robust response.

A second approach to form processing takes advantage of the fact that luminance, oriented bars, corners, and contours can also be understood in terms of spatial correlations. In this framework, the detection of luminance and orientation can be described in terms of detecting 1^{st} and 2^{nd} order spatial correlations, respectively. Corners and contours require a sensitivity of the visual system for three or four points (multipoint) in space (Heydt & Peterhans, 1989; Ito & Komatsu, 2004; Lee & Nguyen, 2001). This suggests that a framework based on multipoint correlations can be fruitful to understand visual form perception.

Within this framework, a library of artificial textures with specific 3rd or 4th and matching 1st and 2nd order spatial correlations was created (Julesz, 1981). Figure 13 shows examples of such textures. With behavioral studies it has been shown that the visual system is sensitive to these artificial textures (Victor & Conte, 1991). More recently, single cell recordings in the anesthetized macaque showed that some V1 and many V2 neurons were selective for the higher order textures (Yu, Schmid, & Victor, 2015). This finding not only further confirms that V2 neurons are tuned to complex visual features, but also challenges current models of early visual form processing such as (Hubel & Wiesel, 1962) that are primarily based on 1st (mean luminance) and 2nd (spatial frequency content) order spatial correlations.

In this section we gave an overview of the current state of the field. In chapter 3 we will show how the ANN approach generates new insights that challenge the feedforward view of early visual form processing. First, we will cover the basic mechanism of ANNs.

1.3 ARTIFICIAL NEURAL NETWORK APPROACH IN SYSTEMS NEUROSCIENCE

1.3.1 Artificial neural networks

Typically, an ANN consists of three or more layers with units (one input layer, one or more hidden layers and one output layer) that are connected with adjustable weights (Figure 3A). The total input of each unit (indexed by *i*) is given by the weighted sum of its inputs plus its bias value: $\mathrm{x_i} = \sum_k^i \mathrm{w_{ik}y_k} + \mathrm{b_i}$, where the index *k* runs over all units that are connected to unit i. This total input is passed through a nonlinear transfer function: $y_i = 1/(1 + e^{-\mathrm{x_i}})$ to generate the scalar output.

Initially, the weights between the units are random; they can have any positive or negative value. Input patterns presented to the input layer, therefore, result in random output patterns in the output layer of the network. During the training phase all weights are adjusted with a learning rule that allows the network to map the input to any desired output. For supervised learning, when the output of the network is fitted to a predefined target, the backpropagation learning rule (Sutton et al., 1988) is most often used. The learning rule adjusts the weights such that (sets of) input patterns (i.e. images of dogs and cats) can be mapped onto (sets of) output patterns (i.e. the value 0 for dogs and 1 for cats). Formally, it can be shown that, given enough hidden units and connections, any input-output relationship can be captured by a feedforward network (Funahashi & Nakamura, 1993; Hornik et al., 1988). In other words, connecting many units with adjustable weights allows the network as a whole to capture arbitrarily complex input-output relationships. Feedforward ANNs where all units have identical intrinsic properties can only capture static input-output relationships such as mapping images of dogs and cats to the labels ‘dog’ and ‘cat’. To capture response dynamics, one can add an extra set of adjustable weights that recurrently connect all units of the hidden layer (Figure 3). It has been shown that such a recurrent network can capture the relationship between any input pattern *sequence* and output pattern *sequence*. In theory, for our dogs and cats example, an artificial recurrent neural network (ARNN) could map *videos* of dogs and cats to a time-varying output that signals the number of dogs or cats in the scene.

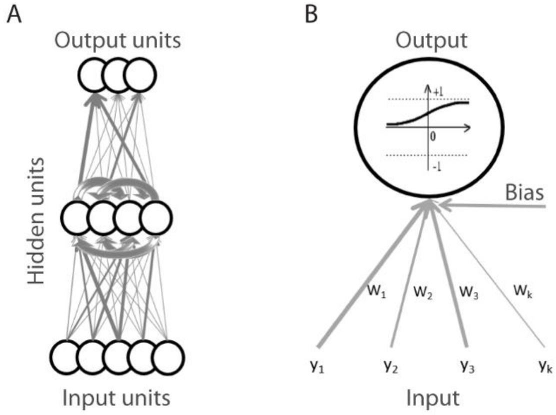

Figure 3. Artificial recurrent neural network. A) A recurrent, three-layer network. All input units are connected to all hidden units, which are in turn connected to all output units. In addition, the units of the hidden layer are all connected to each other; these recurrent connections allow the network to capture temporal relationships between input pattern sequences and output pattern sequences. B) A single unit. A unit in the network receives a weighted sum of its inputs plus a bias value; this sum is passed through a transfer function such as the sigmoid displayed in the inset to generate a scalar output. Adjustment of the input weights during the training phase changes the output for any given set of inputs and gradually allows the network as a whole to capture any desired input-output relationship.

One reason that the field of computational systems neuroscience has not widely adopted ANNs to model brain processes may be that the ANN is only a poor approximation of the biological neural network. First, all units of an ANN process information in an identical way, which is in strong contrast to the large number of neuronal cell types found in the brain. Second, the output weights of a unit can have both positive and negative values, while neurons are excitatory or inhibitory, but never both. Third, in the ANN the weights define connectivity and after the training phase these weights are static. In neurons connectivity is much more complex, dynamic, and affected by ion channel and receptor distributions, neurotransmitter availability, and neuromodulators. Finally, most ANNs have a continuous output (i.e. any value between zero and one for a logarithmic transfer function), whereas neurons typically communicate through all or nothing spikes. In sum, individual units and connections of an ANN do not even remotely resemble individual neurons and their synapses in the human brain.

A more fruitful interpretation of the units, one that sidesteps much of the discussion of biological plausibility, is to assume that the units are equivalent to groups of neurons, or cell assemblies. For instance, cell assemblies can have a net excitatory or inhibitory influence on each other, and due to the potentially large number of neurons in the assembly this influence (the total spike count of all the constituent neurons) is effectively continuous.

Quite separate from the interpretation of the units of the ANN is the decision which level of description is necessary or appropriate to model the relationship between brain and behavior. ANN models start from the view that complex relationships emerge from the distributed interactions among large numbers of simple units (neurons). This view naturally leads to a focus on the connections and how they need to be adjusted to create function. The simplicity of the units mainly follows from the desire to maintain tractability. With more computational power it is possible that more realistic units could be used in the future (general discussion, chapter 5). Furthermore, our modeling approach takes the view that, while much can be learned from studying the variation in the response (e.g. to identical inputs), the mean firing rate to repeated presentations of the same stimulus is a good starting point to link neural function and behavior. The true value of any model lies in the phenomena it explains, the new conceptual insights it generates, and the novel experiments it inspires.

Next, we discuss how ANNs can be applied to experimental data following a three step approach; neural data collection, fitting an ANN to the experimental data, and investigating the properties of the fitted network (ANN approach). Because the ANN approach is substantially different from typical modeling studies, this section is intended as a broad introduction. We will use the MT case study to highlight some of the important experiment and model design choices that also applies to the other two studies we describe in detail in chapter 3 and chapter 4.

1.3.2 Neural data collection

We hypothesized that temporal filtering in motion detection is critically dependent on recurrent connections without the need to invoke other forms of temporal delays such as feedforward temporal delay lines or intrinsic cell properties. To test our hypothesis, we first recorded the velocity response properties of single cells in area MT. We probed the neurons with 14 motion conditions; seven speeds in the preferred and anti-preferred motion direction (see example MT cell Figure 1). Because ANNs are firing rate models and not well suited for single cell spiking responses, we estimated the underlying firing rate for each MT cell by averaging over several trials per motion condition (see error bars in Figure 1). Whereas most analyses and motion modeling attempts ignore the temporal dynamics of velocity tuning (i.e. they average the response over a large time window), we were particularly interested in the emergence of speed tuning and direction selectivity. We hypothesized that the temporal dynamics of the MT cell responses could provide us insights into how the velocity was calculated by the neural system. Therefore, we matched the time scale of the dynamics to the time scale of stimulus presentation. In other words, we defined the main goal of the ANN to learn the relationship between the experimental motion stimuli (monitor frame by monitor frame) and the motion tuned MT response dynamics (monitor frame by monitor frame).

1.3.3 Fit an ANN to the experimental data

Because individual units in an ANN process information at the same time, most feedforward ANNs aren’t capable of fitting sequences of events such as motion signals (see chapter 5 for exceptions). For motion detection, we asked whether the necessary delays can emerge from recurrent connections. There are many ARNNs (see chapter 5), but for the purpose of gaining insights into the relatively short time window in which motion tuning emerges in area MT cells, we fitted our experimental MT data (and the other two data sets we discuss in chapter 3 and chapter 4) with a relatively simple ARNN called the

Elman recurrent neural network (Elman, 1990). We choose this model for its relative simplicity (compared to other more recent and advanced ARNNs) and broad applicability (the recurrent connections guarantee that any input pattern sequence can be mapped onto any output pattern sequence).

Ideally, we would fit the ARNN to the exact inputs and outputs of the experimental data. However, this would require estimating the tuning properties for each MT neuron with a impractically high number of spatial patterns per motion condition. Only a large enough sample of the stimulus-response properties would guarantee a fit of the ARNN to the true underlying velocity tuning properties and not just the speed and direction selectivity for a relatively small subset of spatial patterns. Due to time constraints that are common in experiments with awake behaving animals, we instead took advantage of the pattern invariance property of MT neurons (Albright, 1984). Because most MT cells respond with similar motion tuning properties for many different kinds of spatial patterns, we first estimated the motion tuning properties of the MT cells with a limited number of spatial patterns per motion condition to average out stimulus independent neuronal noise. We then fitted the ARNN to the mean MT response with millions of spatial patterns per motion condition (see Figure 4 for an example input pattern sequence).

It is important to realize that by capitalizing on the assumption of pattern invariance we introduce exactly the kind of biases that one might ideally wish to prevent with a data-driven ANN modeling approach. For instance, it is possible that MT cells are not fully pattern invariant, a feature that an ANN could potentially bring to light if a different experimental design had been chosen (see chapter 5). Practical limitations on data collection or model fitting will often make such tradeoffs necessary.

We were mainly interested whether a network with recurrent connections, but without feedforward temporal delays could reproduce motion tuned response properties. Therefore, we primarily focused on obtaining the best possible fit of the (average) MT response based on millions of motion stimuli. However, general practice is to divide the experimental data into a train set and a test set to prevent overfitting of the data. During the trainings phase the network is only fitted to the train set and the test set is used to measure how well the network generalizes beyond the train set. In other words, this process assures that the network captures general input-output relationships rather than individual examples of specific motion patterns mapped to specific outputs (see chapter 4 where we applied the ANN approach to a data set that did allow for a split into a train set and a test set).

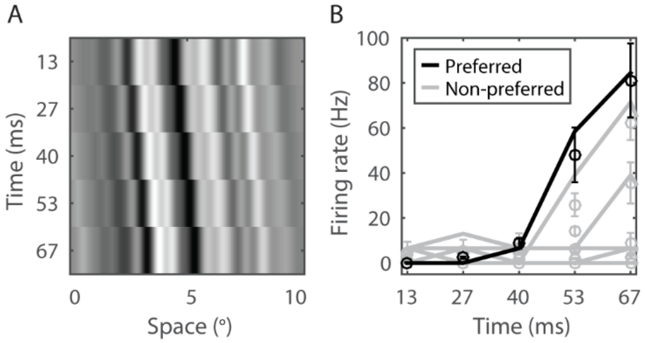

Figure 4. Example motion input and output. A) Example input for the ARNN. Motion stimuli for the ARNN were generated with low pass filtered white noise that moved with one of seven speeds in the preferred or anti-preferred direction (similar to the motion conditions of the experiment). The figure shows an example stimulus for 16 °/s in the preferred direction over five monitor frames. B) Example output for the ARNN. The motion stimulus sequence is fitted to the measured MT cell response for the corresponding motion condition over the five monitor frames (black line, gray lines show the MT response for the six other speeds in the preferred direction). Lines show the experimental data, circles the fitted ARNN response (error bars indicate standard deviation over trials per motion condition).

1.3.4 Investigate the fitted ANN

ANNs provide possible solutions, but because they are unconstrained by anatomical or other knowledge about the neural system, the solution must be verified against neural data, known properties of the neural system, or by further experimentation. In this phase, it is essential to keep in mind that properties of individual network elements will not necessarily map onto the response properties of individual neurons. In other words, they must be treated as qualitative predictions of the properties of neural populations.

The extraction of the ANN mechanism is complicated by the fact that information and computation are inherently distributed across many elements. In other words, one can rarely point at individual network units or weights as being responsible for a specific component of the input-output mapping. This is in strong contrast to many rule-based models where the function of each element is typically well understood. The main challenge of the ANN approach, therefore, becomes to extract network properties that provide insights into the response properties of the constituent neurons and how they perform a specific computation. After training, the ANN is a dynamical system with complex response properties that defy simple descriptions; not unlike the neural system we started out with.

A major advantage of the ANN over a biological neural system, however, is that the artificial network can be studied and modified at will, and this allows one to make use of powerful methods for systems identification that rapidly run into practical feasibility issues when used for biological systems. For instance, white noise reverse correlation (Marmarelis & Marmarelis, 1978), a technique that is commonly used to investigate response properties of neurons in the real brain, proved useful for analyzing the mechanisms of our ARNN.

With the reverse correlation method one can approximate a complex nonlinear input-output relationship with a feedforward model. This method is particularly useful for complex high dimensional data sets such as our MT recordings or the fitted ARNN because it allows for a lower dimensional and more intuitive feedforward description of the functional network properties. The method starts with a simulation of a long sequence of rapidly changing white noise stimuli while measuring the responses of neurons or artificial network elements. In theory, when enough noise stimuli are simulated, all tuning properties of the cell or unit will be part of the stimulus-response set (bounded by the spatiotemporal properties of the noise stimuli). The next step of the ANN approach is to extract the subset that best describes the relationship between stimulus and response as a feedforward process. For example, the tuning of a detector selective for white oriented bars can be extracted with reverse correlation by estimation of the spike triggered average (STA). The feedforward approximation for this detector consists of one space-time filter (the STA) that indicates to what space-time input the detector is most responsive. For more complex detectors (e.g. those that respond both to white and to black bars) techniques such as spike triggered covariance (STC) can be used (Chichilnisky, 2001; Rust, Schwartz, Movshon, & Simoncelli, 2004; Simoncelli, Paninski, Pillow, & Schwartz, 2004).

One of the main challenges with the reverse correlation method is that it suffers from the so called ‘curse of dimensionality’. Generally speaking, more complex input-output relationships require observing the input-output relationship for an impractically high number of noise stimuli. This is especially problematic for experimentalists with strict time constraints that are often forced to intelligently reduce the input-output space such that the tuning properties embedded in a subset are easier to find. Unfortunately, such reductions introduce strong biases that quickly negate the main advantage of the relatively unbiased reverse correlation technique.

For simulated models, however, there are few time constraints, only computational power and memory limitations. This provides the opportunity to reliably approximate feedforward models for the ARNN output units with a simulation of millions of noise patterns combined with the reverse correlation method. As we mentioned previously, we emphasize that ANNs organize themselves in hard to interpret ways and reverse correlation is a valuable tool to help gain insights into how the network solved the input-output mapping and we used this approach throughout the thesis. This is in strong contrast to typical rule based models where design and function of each element is tightly controlled by the experimenter. Here, the reverse correlation technique would most likely not make novel unexpected predictions. In other words, the fundamental difference between the rule based and ANN approach is that the former is driven by formalizing (modeling) informal knowledge (derived from experimental data) while the latter is data analysis where the experimental data (input stimuli and neuronal response) is fully captured in a model that lends itself well for further analyses (such as reverse correlation or perhaps even a rule based model).

Various other techniques can be used to investigate the response properties of the ARNN elements that are practically out of reach for experiments on the real brain. Particularly insightful were exhaustive simulations of the ARNN hidden units that allowed us to systematically search for any kind of tuning property of each of the elements within the network. Combined with a read-out of the adjusted feedforward and recurrent input weights we gained valuable insights into how the network computes motion and form with testable hypotheses about neurons in the motion and form processing pathway (see Chapter 2, 3, and 4).

Ideally, we would test the model predictions on the MT or V2 neurons we used to fit the ANNs. Unfortunately, with single cell recordings, the neurons are lost after each recording session. In chapter 4 we will present the ANN approach that we applied to a data set that did not suffer from this limitation. First, however, we will cover our MT case study in more detail in chapter 2 and our ANN approach applied to the V2 data set in chapter 3.

2 MOTION DETECTION BASED ON RECURRENT NETWORK DYNAMICS

This chapter has been published as: Joukes, J., Hartmann, T. S., & Krekelberg, B. (2014). Motion detection based on recurrent network dynamics. Frontiers in Systems Neuroscience, 8 (December), 239.

2.1 INTRODUCTION

Successful interaction with a dynamic environment requires a neural mechanism for the detection of motion. In the dominant model of motion perception in the primate — the motion energy (ME) model (Adelson & Bergen, 1985; Krekelberg, 2008; Watson & Ahumada, 1985) — the temporal component that is essential for the detection of motion is implemented as a class of neurons that have slow response dynamics. Even though the primate visual system contains a class of slower neurons (the parvocellular stream) the evidence that they are a critical component in motion detection (Colby, 1981; De Valois &

Cottaris, 1998; Nealey & Maunsell, 1994; De Valois et al., 2000) is controversial. For instance, layer IVCα of the primary visual cortex (V1), contains numerous direction selective (DS) cells, but mainly receives magnocellular input (Blasdel & Fitzpatrick, 1984) and, consistent with this, inactivation of the magnocellular layers of the lateral geniculate nucleus (LGN) disrupts motion processing in the middle temporal area (MT), while inactivation of parvocellular layers has little effect (Maunsell & Nealey, 1990). Hence, even though it is clear that motion sensitive neurons receive two sets of inputs, one delayed with respect to the other (De Valois & Cottaris, 1998; Priebe & Ferster, 2005; De Valois et al., 2000), the origin of these delays remains unknown. The model of Maex and Orban (Maex & Orban, 1996) shows that the intrinsic differences between slow (NMDA) and fast (GABA) synaptic transmission could be one source of the necessary delays. However, such intrinsic differences are fixed, and it is difficult to see how they alone can account for the observed wide range of preferred speeds.

Even though anatomically cortical networks are clearly dominated by recurrent connections, this connectivity plays at best a subordinate role in many models of motion detection. For instance, the ME model was originally envisaged as entirely feedforward although it has been extended with recurrent connectivity to amplify direction selectivity (Douglas, Koch, Mahowald, Martin, & Suarez, 1995; Maex & Orban, 1996; Suarez, Koch, & Douglas, 1995) or motion integration and segmentation (Bayerl & Neumann, 2004; Tlapale, Masson, & Kornprobst, 2010). The analytic work of (Mineiro & Zipser, 1998; Sabatini & Solari, 1999), however, has shown that recurrent connectivity alone is in principle sufficient to generate direction selectivity and (Clifford, Ibbotson, & Langley, 1997; Clifford & Langley, 2000) mathematically showed that a recursive implementation of the temporal filter of the ME model can greatly reduce the amount of storage and computation needed for a motion detector tuned to a broad spatiotemporal frequency range.

Our works starts from the data – a set of recordings from MT neurons – and shows that an artificial recurrent neural network can faithfully reproduce the speed and direction tuned responses to visual motion. New insights into motion mechanisms resulted from a detailed, quantitative investigation of this network. Notably, no separate classes of fast and slow neurons, or carefully tuned delay lines were needed to generate a wide range of speed preferences. Instead, a range of temporal delays and concomitant speed preferences emerged from the weight patterns of the network. Second, while the recurrent network could be approximated by a ME model, such a feedforward approximation failed to capture the sequential recruitment typically found in MT neurons (Mikami, 1992). Finally, the response properties of the units in the recurrent network (e.g. Gabor receptive fields, simple- and complex-like responses), showed a remarkable match with the known properties of neurons in the motion processing pathway.

2.2 MATERIALS AND METHODS

2.2.1 Experimental data

2.2.1.1 Subjects

We measured the speed tuning properties in area MT of two adult male rhesus monkeys (Macaca mulatta). Experimental and surgical protocols conformed to United States Department of Agriculture regulations and the National Institutes of Health guidelines for humane care and use of laboratory animals and were approved by the local IACUC committee.

2.2.1.2 Visual stimulation

The visual stimuli were generated with in-house OpenGL software (Quadro Pro Graphics card, 1024x768 pixels, 8 bits/pixel) and displayed on a 21 inch monitor (75 Hz, non-interlaced, 1024x768 pixels; model GDM-2000TC; Sony). Monkeys viewed the stimuli from a distance of 57 cm in a dark room (<0.5 cd/m2) while seated in a standard primate chair (Crist Instruments, Germantown, MD) with the head post supported by the chair frame. We sampled eye position at 60 Hz using an infrared system (IScan, Burlington, MA), and monitored and recorded the eye position data with the CORTEX program (Laboratory of Neuropsychology, National Institute of Mental Health, Bethesda, MD; http://www.cortex.salk.edu/), which was also used to implement the behavioral paradigm and to control stimulus presentation.

2.2.1.3 Stimuli and experimental paradigm

We mapped velocity tuning with a random dot pattern that consisted of 100 dots within a 10° diameter circular aperture. The dots had infinite lifetime and were randomly repositioned after leaving the aperture. The dots were 0.15° in diameter and had a luminance of 30 cd/m2. Compared with the 5 cd/m2 background, this resulted in a Michelson point contrast of 70%.

The activity of single units in area MT was recorded with tungsten microelectrodes (3–5 MOhm; Frederick Haer Company, Bowdoinham, ME), which we inserted using a hydraulic micropositioner (model 650; David Kopf Instruments, Tujunga, CA). We filtered, sorted, and stored the signals using the Plexon (Dallas, TX) system. Area MT was identified by its high proportion of cells with directional selective responses, small receptive fields (RFs) relative to those of neighboring medial superior temporal area, and its location on the posterior bank of the superior temporal sulcus. The typical recording depth was in agreement with the expected anatomical location of MT determined by structural magnetic resonance scans.

We determined the directional selectivity and RFs of the cells using automated methods (for details, see (Krekelberg, Vatakis, & Kourtzi, 2005)). Based on the RF center and the preferred direction of motion (rounded to the nearest multiple of 45°) estimated by these methods we optimized stimuli for subsequent measurements. The mean RF eccentricity and SD was 8 ± 4.3° (range of 3 to 15°). The random dot patterns appeared 250ms after the monkey started fixating on a central red dot. After moving in the preferred or anti-preferred direction of the neuron for 500ms, the pattern was extinguished. The range of speeds was 1, 2, 4, 8, 16, 32, and 64°/s. The 14 conditions (7 speeds, 2 directions) were randomly interleaved and repeated between 4 and 21 times. Trials in which eye position deviated from a 2° wide square window centered on the fixation spot were excluded from analysis.

The MT response to the moving stimuli was binned in 13ms time windows; the frame rate of the monitor used during the experiments (75 Hz). This allowed us to investigate the emergence of the speed tuning and direction selectivity properties at a temporal resolution that matched the (apparent) motion on the monitor.

2.2.2 Recurrent motion model

2.2.2.1 Elman recurrent neural network

We modeled the neuronal data with an Elman recurrent neural network (Elman, 1990) implemented in the Matlab Neural Network Toolbox (version 4.0.1). The network consisted of units that are considered a crude approximation of a neuron or a group of neurons (Figure 6A). The units were interconnected with adjustable weights simulating synaptic connections with variable strength. Each unit also had an adjustable bias value.

The network had an input, hidden, and output layer. The input layer consisted of 750 units that simulated a RF of 10° (0.013° per unit); the diameter of the stimulus used during the experiments. The input layer was fully connected to the hidden layer in a feedforward manner. The hidden layer had 300 units that were fully connected to the output layer in a feedforward manner. In addition, all hidden units were laterally/recurrently connected to all hidden units. The output layer consisted of 26 units, each simulating one MT cell. The output for each unit of all layers (indexed by i) was calculated by first determining the weighted sum of its inputs plus the bias value: $\mathrm{x_i} = \sum_k^i \mathrm{w_{ik}} \mathrm{y_k} + \mathrm{b_i}$, where the index k runs over all units that are connected to unit i, and then passing this through a sigmoid transfer function: $y_i = \ 1/(1 + e^{-\mathrm{x_i}})$.

2.2.2.2 Output patterns

We used the model to capture the responses of a representative subset of MT neurons from our sample of 129. To reduce computational complexity, we focused our analyses and modeling on 26 MT cells with robust, direction selective responses, and band pass speed tuning. The specific criteria for inclusion were the robustness of the response (firing rate > 7 spikes/s averaged over all speeds in the preferred direction), modest to strong direction selectivity (DSI>0.1, for definition see below), and a preferred speed in the range of 8-32°/s. This selection resulted in a population of 26 MT cells.

The population response revealed an initial response latency of approximately 30ms followed by the rapid onset of speed and direction tuning that lasted around 100ms, and finally a sustained phase with relatively constant responses and tuning (Figure 5, thin lines). We chose to train the network on the generation of speed tuning and direction selectivity only (27-93ms; Figure 5, thick lines). Given the temporal binning in 13ms time bins (the duration of a monitor frame in the experiment), this resulted in output pattern sequences of firing rates at five time points for each of the 14 conditions (7 speeds, 2 directions) and each of the 26 output units. We normalized the response to a suitable range for the network with a division by the maximum firing rate over all time bins, speeds, directions and MT cells.

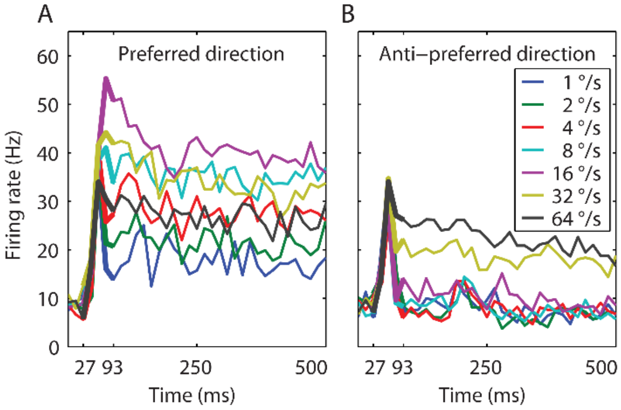

Figure 5. Experimental data. A) The temporal dynamics of the MT population to seven speeds in the preferred direction. The figure shows the average response of 26 MT neurons. The thin lines represent the response to the total duration of the stimulus, the thick lines the time window we used to train the network. An initial transient lasted 67ms during which speed tuning and direction selectivity started to emerge. Speed tuning was maximal around 93ms and followed by a slow reduction in firing rate for most speeds. B) The temporal dynamics of the MT population to seven speeds in the anti-preferred motion direction. Here too, the initial transient lasted 67ms and direction selectivity was maximal around 93ms, followed by slow adaptation. The onset response was not strongly direction selective, but after 93ms the response to the preferred direction could be twice as large as the response

to the anti-preferred direction. These data document that the response, speed tuning, and direction tuning change dynamically in the first 100ms after stimulus presentation.

2.2.2.3 Input patterns

We recorded responses to preferred and anti-preferred directions of motion only and, therefore, did not attempt to model the entire two-dimensional random dot patterns. Instead, we represented the input as binary random dot patterns and trained the network to respond in a tuned manner to each of these patterns (Figure 5B). To create the input patterns, 750 (the number of input units) values were randomly assigned a negative (black) or positive (white) constant chosen to ensure that the final input values were almost all (4 SD) between -1 and 1. These binary noise values were spatially low pass filtered by convolving with a Gaussian ($\sigma = 0.25°$) and a multiplication with a Gaussian envelope over the whole input space ($\sigma = 2.5°$) to reflect the spatial limits of the RF.

A moving input pattern sequence was modeled by shifting the input pattern in the preferred or anti-preferred direction with one of seven speeds. In the physiological experiments, the visual pattern moved between 0.013° and 0.85° per monitor frame (1°/s to 64°/s, respectively). In the model this was implemented by shifting the input pattern by 1 to 64 input units per 13ms, respectively.

2.2.2.4 Training phase

Before training the network, we initialized the weights and bias values of all layers with the Nguyen-Widrow algorithm. We trained the recurrent neural network on the input and output pattern sequences we described above in the following way. First, we randomly chose one of seven speeds and a direction of motion. Second, frame-by-frame, a new input pattern sequence for that speed and direction was presented on the input units. Third, for each frame, we calculated the response of the hidden units based on the current feedforward input and the recurrent feedback, and then calculated the response of the output units.

Fourth, the error of the network was defined as the difference between the response of the output units and the response of all 26 MT cells (for that speed and direction, and in the corresponding time bin after stimulus onset). This error was used to modify all connection weights in the network using error back-propagation-through-time. We repeated these steps (epochs) five million times until the network converged to reproduce the response of all 26 MT cells. Network parameters were then frozen and we investigated the trained network.

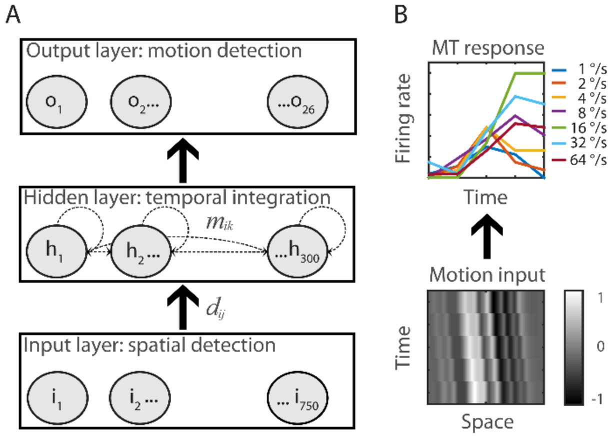

Figure 6. Recurrent motion model. A) An Elman recurrent neural network with 750 input units that were all-to-all connected to the units of the hidden layer. The hidden layer had 300 all-to-all recurrently and laterally connected hidden units (dashed lines) that were all-to-all connected to the 26 output units of the RMM. Adjustable weights allowed the network to map motion inputs onto the speed tuned and direction selective response of the 26 recorded MT cells. B) Example input-output. Low pass filtered binary random dot patterns (range -1 to 1) shifted with 16 °/s in the preferred motion direction over 5 time bins (bottom) is fitted to the measured response of the example MT cell to the preferred speed in the preferred motion direction (top, black line, gray lines represent non-preferred speeds).

2.2.2.5 Reverse correlation

We investigated the recurrent motion model (RMM) as if it were a linear-nonlinear model (LN-model) (Chichilnisky, 2001; Rust et al., 2004; Simoncelli et al., 2004). The LN-model processes information in a strictly feedforward way and is described by a set of linear space-time filters and their corresponding nonlinearities. We used the spike triggered average (STA) and the spike triggered covariance (STC) methods to estimate the filters (Chichilnisky, 2001; Rust et al., 2004; Simoncelli et al., 2004). As mentioned previously, the MT cell response was normalized to create suitable targets for the RMM. To generate the spikes necessary for reverse correlation, we multiplied the activation of each output unit with a constant that generated a maximum of 30 spikes per time bin, followed by a simple rounding to the nearest integer. As noise inputs we used binary random dot patterns identical to a single frame of the moving spatial patterns described previously. It was not feasible to determine the filters for 750 dimensional noise patterns for all of the output and hidden units. We therefore chose stimuli consisting of 0.04° wide bars, reducing the spatial dimension by a factor of three. The reverse correlation history - the number of time bins leading up to the output activity – was set to be 67ms, the time needed for the MT population to create a stable speed tuned and direction selective output. Two million noise stimuli were used to determine the filters for the output units and one million for the hidden units.

We calculated the filter dimensions that contained the most information in the space spanned by both the STA and STC with an information-theoretic generalization of spike-triggered average and covariance analyses (iSTAC) (Pillow & Simoncelli, 2006). We measured the statistical differences between the spike-triggered (the noise inputs that triggered spikes in the output and the hidden units of the RMM) and the raw stimulus ensemble (the noise inputs that were used for the STA and STC) with the Kullback-Leibler divergence, a measure of the difference between two probability distributions (Cover & Thomas, 2012). Finally, we determined the nonlinearity associated with each filter by projecting the raw stimulus ensemble and the spike-triggered stimulus ensemble onto the filters, and dividing the histogram of the projected spike triggered ensemble by the histogram of the projected raw stimulus ensemble, over four standard deviations away from the mean. These nonlinearities are shown in Figure 8E. The four-dimensional nonlinearities were estimated analogously; as the ratio of the four-dimensional histograms of the projections of the spike triggered and raw stimulus ensembles.

We estimated the speed tuning and direction selectivity properties of the LN-model with 1000 new moving input pattern sequences. For each pattern and for each time bin, the inner product between the motion inputs and the filters gave us the projection value. Ideally, passing the projections through a high-dimensional nonlinearity would give us the firing rate of the combined filter model. Due to computer memory constraints, however, we had to restrict this to 1-dimensional nonlinearities; i.e. we assumed that the filter dimensions were separable. The firing rate for the combined filter output was estimated with a simple summation of the individual filter outputs followed by subtracting out the mean for n-1 filters. To determine if this linear summation was a good estimate, we estimated 4-dimensional nonlinearities for the three filter quadruples (two excitatory and two suppressive filters per spatial frequency, see Figure 8B) and compared the predicted firing rate by passing each input pattern through the 4-dimensional nonlinearity (NL) with the linear summation of the 4 individual filters per quadruple (see Discussion).

2.2.2.6 Speed tuning and direction selectivity

For the RMM output and hidden units, the direction selectivity and speed tuning was based on the mean response over time to 1000 new motion inputs for all speeds and directions. For the MT cells we used the mean response over time to all experimental trials. We calculated the direction selectivity index (DSI) with the maximum (average) response over the seven speeds in the preferred direction and the (average) response at that speed in the anti-preferred direction: (preferred – anti-preferred) / (preferred + anti-preferred). For the speed tuning index (SI) we used the maximum and the minimum response across the seven speeds in the preferred direction: (maximum – minimum) / (maximum + minimum).

2.2.2.7 Relative response modulation

We classified the hidden units of the RMM as simple or complex based on their response to sinusoidal gratings with the preferred spatial frequency (0.5 cycles/°) and speed (16°/s) of the network and optimized for motion direction per hidden unit. After presenting the gratings for 10 seconds, we removed the response to the first 67ms (initial transient), and then determined the relative modulation (F1/F0) as the ratio of the response at the grating temporal frequency (F1) to the mean response (F0), averaged over time. The hidden units were classified as simple when F1/F0 > 1, complex otherwise (Movshon, Thompson, & Tolhurst, 1978; Skottun, Valois, & Grosof, 1991).

2.2.2.8 Direct and indirect input

The weights between all input units and the hidden unit defined the hidden unit's direct input. In a feedforward model, the RF would correspond to those input units where the weights are strong enough to drive the hidden unit. In a recurrent model, however, the units are also modulated by input that travels via one or more other hidden units - the indirect input. As a first approximation of this indirect input, we considered only the indirect input that arrives 13ms (one simulation time step) later than the direct input. To quantify the indirect input that hidden unit A receives via all other hidden units, we multiplied the connection strength between hidden unit B and A with the direct input of unit B, and summed this over all hidden units B.

Cross-correlation of the low-pass filtered direct and indirect input provided us with an estimate of the spatial shift (dx) between them. The low–pass filter was identical to the one used for the motion inputs. Because some units had no clear weight patterns, we included only units where the cross-correlation was higher than 0.6 (263 out of 300 hidden units).

2.3 RESULTS

2.3.1 Experimental data

We used the velocity tuning curves of 26 MT neurons recorded in two awake macaque monkeys. The population response to seven speeds in both the preferred and anti-preferred motion direction had an onset delay of 30ms and an initial transient lasting 70ms during which speed tuning and direction selectivity started to emerge (Figure 5). Speed tuning and direction selectivity were maximal and stable 90ms after stimulus onset; although there was a slight overall reduction in firing rate over the remaining 500ms recorded data. This was likely an effect of adaptation, as has previously been reported in MT (Kohn & Movshon, 2003; Krekelberg, van Wezel, & Albright, 2006; Schlack, Krekelberg, & Albright, 2007).

2.3.2 Model training

We first investigated whether a network consisting of (artificial) neurons, all with identical intrinsic properties but modifiable synaptic strengths, could reproduce the temporal dynamics and polarity insensitive velocity tuning of the 26 MT cells. Second, probing this network allowed us to determine how its constituent units and connections solved the complex task of motion processing and how its properties relate to neurons in the motion processing pathway.

We created a recurrent neural network with 750 input units, 300 recurrently connected hidden units, and 26 output units (Figure 6A, see Materials and methods). The visual input patterns were modeled as one-dimensional random dot patterns moving leftward or rightward at one of seven speeds (Figure 6B, example input). The output units were then trained using back-propagation-through-time to reproduce the response of the 26 MT cells when presented with any of the input patterns (see Materials and Materials and methods). In the remainder of the results section we highlight salient properties of the model. First, we show that the network reproduced the MT responses; this is a proof of principle that a recurrently connected network whose units all have identical intrinsic properties could underlie the MT responses. Second, we probed the output units with the reverse correlation methods commonly used in electrophysiology. This will demonstrate that the output units behave very much like a motion energy (ME) detector, while also showing that such a feedforward description does not capture the typical time course of direction and speed selectivity. Third, we investigate how the range of speed tuning in the output units is created from the hidden units. Finally, we investigate the properties of the hidden units to reveal the connectivity from which temporal delays and spatial offsets emerge and result in the computation of the speed and direction of motion.

2.3.3 Proof of principle

The performance of the recurrent motion model (RMM) was tested with a simulation of 1000 new input patterns moving with seven speeds in the preferred and anti-preferred motion direction. The average response over time compares the speed tuning properties of the MT and the RMM population response (Figure 7A). The average over speed compares the temporal dynamics over the trained time bins (Figure 7B, thick lines) and over the full 500ms stimulus presentation time of the physiological experiments (Figure 7B, thin lines). To determine the performance of the individual output units of the RMM, we calculated the speed tuning index (Figure 7C) and direction selectivity index (Figure 7D) for both the MT cells and the output units of the RMM over the trained number of time bins (see Materials and methods). As Figure 7 shows, the RMM captured the speed tuning and direction selectivity properties of the individual MT cells over the trained number of time bins and generalized well (i.e. remained in the correct stable state) to the presentation of motion stimuli for a longer period of time, with low variability across trials (i.e. patterns).

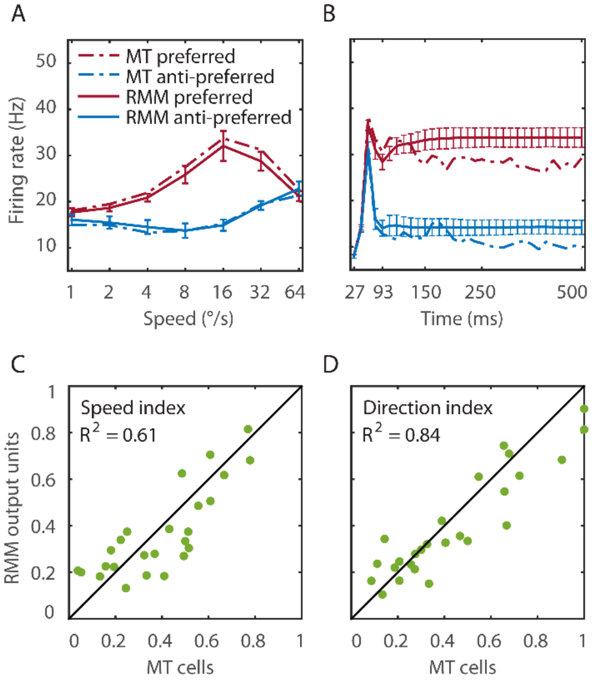

Figure 7. Speed tuning and direction selectivity of the MT cells and RMM output units. We simulated the response of the RMM output units to 1000 new patterns and compared their response to the measured MT response. A) The average response over the first 93ms after stimulus onset. The response is shown for each of seven speeds in the preferred (black) and anti-preferred (gray) motion direction. The dotted lines represent the MT population response, the solid lines the population output of the RMM. The error bars indicate 1 standard deviation over trials. B) The time course of the response averaged over the seven speeds for the trained number of time bins (thick lines) and for the full 500ms of the

physiological experiment (thin lines). C-D) The speed tuning index (C) and direction selectivity index (D) for the output units (x-axis) and the MT cells (y-axis). This figure shows that the RMM faithfully captured the temporal dynamics (except for the short term adaptation) as well as the speed tuning and direction selectivity of both the MT population and the single cell response.

While the generalization to a new set of random dot patterns demonstrates a degree of robustness and pattern invariance of motion detection in the RMM, a more stringent test is to consider directional selectivity for patterns that were qualitatively different from the (random dot) patterns used in the training procedure. We determined the velocity curves in response to drifting sine wave gratings (SF = 0.5°), and found that these were highly correlated with the velocity curves measured with random dot patterns (R^2=0.95). This shows that the RMM was a robust motion detector with a high degree of pattern invariance, consistent with the known properties of area MT (Albright, 1984).

2.3.4 Comparison with the Motion Energy Model

A common problem in neural network modeling is that one can rarely point at individual elements of the network model as being responsible for a specific component of the input-output transformation. The reason for this is that information and computation are inherently distributed across many elements. This is the same problem experimentalists face when they investigate the motion processing pathway in the real brain. One approach that provides a lower-dimensional description of a complex system uses noise stimuli together with reverse correlation analysis (Chichilnisky, 2001; Rust et al., 2004; Simoncelli et al., 2004). This technique describes a neuron in terms of an equivalent feedforward linear non-linear model by estimating a set of linear filters and their static nonlinearities (LN-model, see Materials and methods). We used this method here to gain insight into the RMM and to allow a direct comparison with known LN-models such as the ME model and LN-models based on the response of real neurons to noise stimuli. We presented visual noise to the RMM and performed an information theoretic spike triggered average and covariance analysis (iSTAC) (Pillow & Simoncelli, 2006) to estimate the most informative filters. For each filter we also calculated the nonlinearity (Chichilnisky, 2001). Figure 8A shows the estimated LN-model of one of the output units. The spike triggered average (STA, Figure 8B) did not show any clear slanted space time structure nor did it contain much information (STA, Figure 8E). This is expected since the output units were trained to be polarity insensitive, hence little information should be contained in the STA. The 13 most-informative iSTAC filters (Figure 8E), however, were clearly slanted in space-time. Six filters were tuned to the preferred direction and six filters were tuned to the anti-preferred direction (Figure 8C). Stimuli that matched a filter (high positive axis projection in Figure 8F) or that matched the polarity inverse of a filter (large negative axis projection in Figure 8F) evoked similar responses. In other words, the nonlinearities were symmetric, hence the output of the filters was polarity insensitive. For the excitatory filters, the unit’s response increased with the match between stimulus and filter. The opposite was true for the anti-preferred filters; the greater the match with the filter, the smaller the response of the output unit. These filters were suppressive. All pairs of preferred and anti-preferred filters were phase-shifted with respect to each other.

If MT neurons were perfect ME detectors, one would predict two pairs of oriented, phase shifted space-time filters, followed by quadratic nonlinearities that evoke excitation from the preferred direction of motion and inhibition from the anti-preferred direction of motion (i.e. opponency). The properties of the first pair of excitatory and the first pair of suppressive filters (first quadruple) qualitatively matched this prediction. However, reverse correlation of this output unit revealed two additional quadruples of filters sensitive to higher spatial frequencies, but overlapping in space-time (Figure 8D). We also note that – unlike the prediction of the ME model – the preferred and anti-preferred filters were not perfectly mirror symmetric. This can be seen most easily in the Fourier spectra (Figure 8C, alongside the filters) and the pooled excitatory and suppressive spectra (Figure 8D). Energy in the preferred filters was concentrated around the preferred speed and direction. The spectrum of the anti-preferred filters, however, was more widely distributed over temporal frequencies (see Discussion).

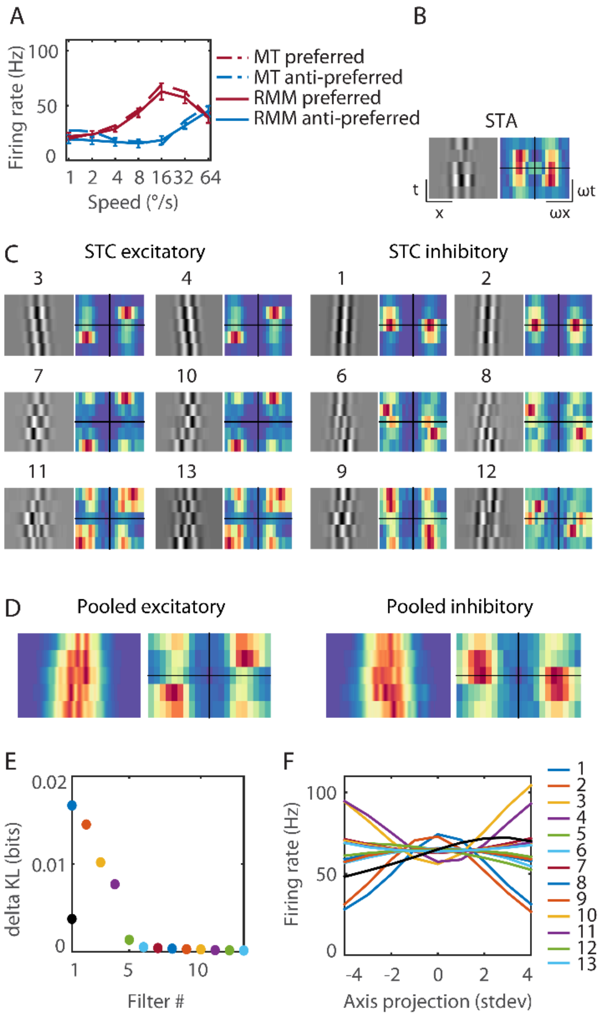

Figure 8. Reverse correlation analysis of the RMM output units. A) The response at 93ms after stimulus onset for an example MT cell/output unit. The response is shown for each of seven speeds in the preferred (black) and anti-preferred (gray) motion direction. The dotted lines represent the example MT single cell response, the solid lines the example output unit of the RMM. The error bars indicate 1 standard deviation over trials. B-C) The 13 filters of the RMM ordered by the amount of information (numbers above the filters show rank order). The STA displayed in B) had no clear slanted space time structure as expected for a polarity insensitive output unit. The six excitatory filters (Excitatory) had a

rightward slant; they were tuned to the preferred speed and direction, with increasing spatial frequencies per quadrature pair. The six suppressive filters (Suppressive) had a leftward slant; they were tuned to the anti-preferred direction, with increasing spatial frequencies per quadrature pair. The Fourier spectra are shown next to the 13 filters. For the excitatory filters power was concentrated around the preferred speed and direction of the output unit (16 °/s). The Fourier spectra of the suppressive filters had a wider distribution of power over temporal frequency, ranging from stationary to the preferred temporal frequency in the anti-preferred direction and to fast speeds in the preferred direction. D) Pooled excitatory and suppressive filters and Fourier spectra. The pooled excitatory and suppressive filters show that the individual filters largely overlap. The pooled excitatory and suppressive Fourier spectra exemplify the asymmetry of excitation and suppression. E) The amount of information (in bits) contained within the STA (black), the excitatory filters (red) and suppressive filters (blue). (F) The nonlinearities of the 13 filters. Stimuli that matched a filter (high positive axis projection) or that matched the polarity inverse (high negative axis projection) resulted in a high firing rate for the excitatory filters (red) and a low firing rate for the suppressive filters (blue).

The asymmetric spectra from the 26 output units grouped by speed preferences of 8, 16, and 32 °/s are shown in Figure 9. As expected, the Fourier spectra of the STA contained almost no motion energy. The pooled excitatory spectra of the iSTAC filters where sharply tuned to the preferred speed, while the pooled suppressive spectra were relatively broad. Unlike the prediction of the ME model, the spectra were not mirror copies of each other, which shows that the RMM does not perform a strict subtraction of opposing directions of motion (i.e. motion opponency), but a broader suppressive interaction among multiple Fourier components. We emphasize here that the asymmetry of the filters is a feature of the MT data that was successfully captured by the RMM.

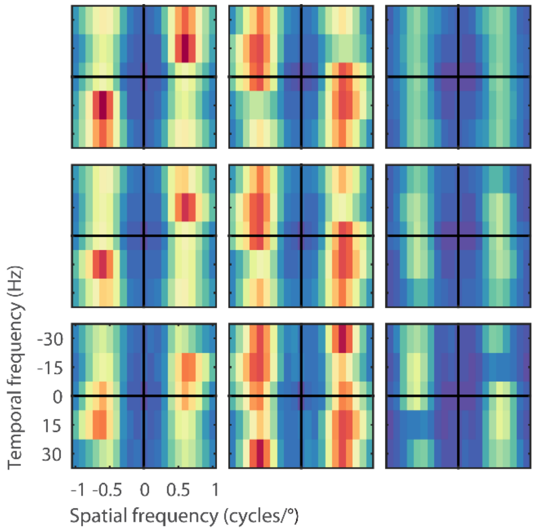

Figure 9. Velocity computation with asymmetric excitation and suppression. Pooled excitatory, suppressive, and STA Fourier spectra of the filters (from left to right) for the 26 output units grouped by their preferred speed of 8, 16, and 32°/s (from bottom to top), normalized to the peak power per unit. Excitation was always centered on the preferred speed and direction of the unit. Suppression was more broadly distributed across temporal frequencies. As expected, the STA had relatively little motion energy.

2.3.5 Limitations of feedforward models

Reverse correlation analysis allowed us to reduce the high-dimensional description of the RMM to 13 linear filters and static nonlinearities per output unit. In other words, we determined LN-models that closely matched the input-output relationship of each of the RMM output units. We can now investigate the extent to which these LN-models capture the full motion response of the MT data that was closely matched by the RMM. We simulated 1000 motion inputs with seven speeds in both directions and presented them to both the LN-models and the RMM.

Figure 10 shows the time course of direction selectivity (A) and speed tuning (B) for an example MT neuron (dotted curve) over the first 4 motion steps (time bins 2-5). Both the DSI and SI rise rapidly over the course of the first few time bins of stimulus presentation. Such sequential recruitment has been reported before (Mikami, 1992)f. As previously shown in Figure 7, the recurrent model captures this nonlinear behavior quite accurately over the trained number of time bins (solid curve). The dashed curve shows the time course of the LN model -the best second-order feedforward approximation to the RMM- for this unit. Clearly, the LN model underestimates the nonlinear response properties. Panels (C) and (D) in Figure 6 confirm that this is a consistent finding across the sample of 26 output units. For each unit we calculated the DSI and SI per motion step for the MT neuron, and the corresponding RMM and LN model units. Figure 10 compares the average DSI (C) and SI (D) across the four motion steps between the two models and the MT data. Whereas the output units of the RMM captured most of the speed tuning and direction selectivity properties of the MT cells (R^2=0.6 (DSI), R^2=0.8 (SI)), the corresponding LN-models performed much worse (R2=0.3 (DSI) and R2=0.5 (SI)). The mismatch between the feedforward model and the MT units was particularly large for MT cells with high direction selectivity and speed tuning.

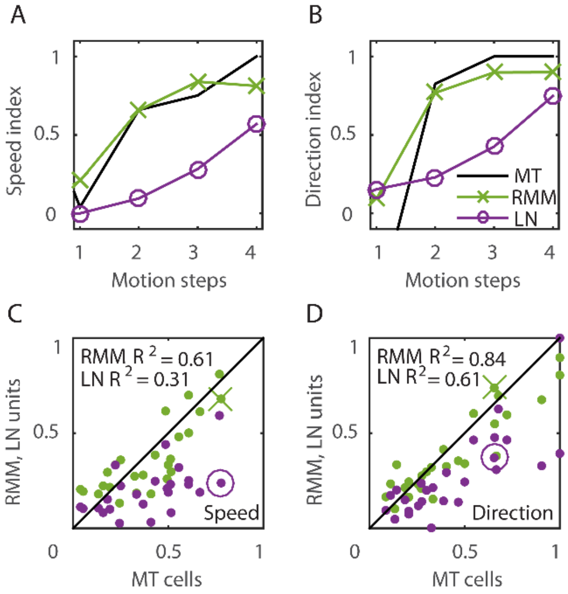

Figure 10. Temporal dynamics. A-B) Speed tuning index (A) and direction selectivity index (B) as a function of the number of motion steps for an example MT cell (dotted square) and corresponding RMM output unit (dashed circle) and LN-model output (solid cross). C-D) Speed tuning index (C) and direction selectivity index (D), averaged over motion steps, for all RMM output units (circles) and their LN-models (crosses) as a function of MT cells. This figure shows that, while the output units of the RMM captured the full speed tuning and direction selectivity properties of most MT cells, the LN model generally fails to do so.

This finding strongly suggests that an LN-model based on first (STA) and second (STC) order space-time correlations is not sufficient to explain the response of single MT cells. Making this claim we need to address two issues. First, the LN-models are based only on the 13 most-informative dimensions; the full spike triggered covariance contains thousands more dimensions that could collectively describe a considerable amount of information. To address this issue, we note that an LN-model based on filters beyond the 13th filter (up to 100 tested) did not improve speed tuning and direction selectivity compared to the 13 filter LN-model (data not shown). This strongly suggests that the mismatch between the LN model and the data is not due to second-order filters that we excluded from the LN model. Second, for all LN-models, we summed the output of the individual filters to determine the combined filter output (see Materials and methods) and it is possible that the individual filters should be combined nonlinearly to accurately describe the velocity response for the LN-models (Chichilnisky, 2001; Rust et al., 2004; Simoncelli et al., 2004). The curse of dimensionality, however, prevents us from accurately estimating the full 13-dimensional nonlinearity. We did, however, estimate 4-dimensional nonlinearities for each of the three filter quadruples with reasonable accuracy (see Materials and methods). Qualitatively, visual inspection of the 4-dimensional nonlinearities did not reveal interactions. This suggests that our separable approximation of the high-dimensional nonlinearity was appropriate. Quantitatively, passing the filter outputs through the 4-dimensional nonlinearities did not improve speed tuning or direction selectivity compared to the separable combination of the filters (data not shown). We also note that the sharp reduction in the amount of information contained in each filter quadruple for increasing spatial frequencies matched a sharp reduction in speed tuning for higher spatial frequency quadruples. For instance, even though the amount of information contained in the third quadruple was not negligible, the contribution to the overall mean speed tuning and direction selectivity was very small. This is a strong indication that filters beyond the 13th filter do not contribute to the speed tuning and direction selectivity of the LN-models, and that the differences between the LN-model on the one hand and the RMM output unit and the MT cell response on the other, are due to third and higher order space-time interactions.

2.3.6 Speed preferences in the output units

The RMM generates output units with a range of preferred speeds that matches our sample of MT neurons. The model allows us to investigate how this range is constructed from a single population of hidden units. Figure 11 shows the strength of the connection between neurons in the hidden layer and those in the output layer. In panel (A) we sorted both the hidden units (x-axis) and the output units (y-axis) according to their preferred speed for motion in the preferred direction. This reveals, not unexpectedly, that output units with, for instance, high preferred speeds had excitatory connections to hidden units with matching high preferred speeds, and inhibitory connections to units with non-matching preferred speeds. Panel (B) shows the same connection strengths, but now we sorted the hidden units (x-axis) according to the preferred speed for motion in the anti-preferred direction. This shows that many output units have relatively strong connections to neurons that prefer fast speeds in the anti-preferred direction and are inhibited by neurons that prefer low and intermediate speeds in the anti-preferred direction.

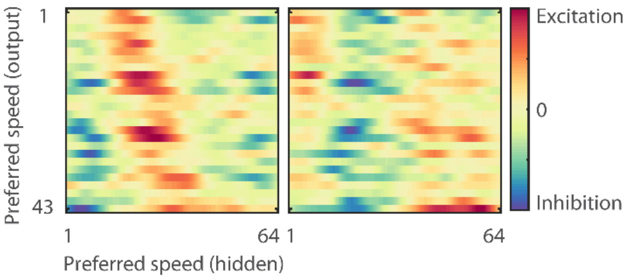

Figure 11. Weights of the hidden to output units. A) Strength of the connection between neurons in the hidden layer and those in the output layer sorted according to the preferred speed of the hidden units (x-axis) and the output units (y-axis) in the preferred motion direction. This panel shows, for instance, that the output units that preferred high speeds had excitatory connections to hidden units with high preferred speeds, and inhibitory connections to units with low preferred speeds. B) Strength of the connection between neurons in the hidden layer and those in the output layer sorted according to the preferred speed of the hidden units in the anti-preferred motion direction (x-axis) and the preferred speed of the output units in the preferred motion direction (y-axis). This panel shows that many output units have relatively strong connections to neurons that prefer fast speeds in the anti-preferred direction and are inhibited by neurons that prefer low and intermediate speeds in the anti-preferred direction.

This connectivity analysis shows that the motion tuning of the output units arises from the weighted combination of the motion tuning of the hidden layers. They receive excitation from hidden units with matching preferred speeds in the preferred direction as well as from hidden units with non-matching preferred speeds in the anti-preferred direction. And, they are inhibited by hidden units with non-matching preferred speeds in the preferred direction as well as by hidden units with matching preferred speeds in the anti-preferred direction. While this provides an intuitive explanation of the motion tuning of the output units, it obviously raises the question how the hidden units generate their motion tuning. We turn to this question next.

2.3.7 Hidden units: tuning properties

We determined direction and speed preference for all hidden units by presenting moving patterns (see Materials and methods). The preferred speeds of the hidden units ranged from 1 °/s to 64 °/s in both the preferred (143 units, mean 30 °/s) and anti-preferred (157 units, mean 30 °/s) direction of the RMM population response (Figure 12A). Many of the hidden units that were tuned to the preferred direction also preferred the same speed as the RMM population. In contrast, many of the hidden units that were tuned to the anti-preferred direction preferred speeds lower than the average preferred speed of the RMM population.

Second, to classify hidden units as simple or complex, we presented sinusoidal gratings, recorded the responses, and determined the ratio of the response modulation at the temporal frequency of the stimulus and the mean response. This is the F1/F0 ratio - a measure often used to categorize simple and complex cells (Dean & Tolhurst, 1983; Movshon et al., 1978; Skottun et al., 1991). Across the population of hidden units, the distribution of F1/F0 was not bimodal (Figure 8C), suggesting a continuum of simple and complex units, but we used the conventional cutoff (Skottun et al., 1991) of F1/F0>1 (see Materials and methods) to classify 124 hidden units as simple-like (black bars) and 176 hidden units as complex-like (gray bars).

Third, we investigated how the hidden units were connected to the input and to other hidden units. Many hidden units had Gabor-like input weights, but others had no clear spatial structure. Principal component analysis on the input weights of all the hidden units revealed six components that explained 98% of the variance (data not shown). These six components could be grouped into three pairs whose input weights were roughly in quadrature, but with increasing spatial frequencies. These weight patterns provide the building blocks that lead to the filters that were extracted with reverse correlation of the output units (Figure 8).

2.3.8 Hidden units: noise analysis

We used the same white noise analysis previously applied to the output units to gain more insight into the functional properties of the hidden units. First, both simple-like and complex-like units had multiple slanted filters with excitatory and/or suppressive symmetric nonlinearities. Excitatory filters typically corresponded to the preferred speed and direction while suppressive filters corresponded to the anti-preferred speed and direction of the unit, albeit with broader tuning. In other words, their properties were qualitatively similar to those of the output units (Figure 9). Second, the amount of information in the STA for each hidden unit correlated well with the F1/F0 ratio (r=0.78, p<0.001), confirming that simple units had STA's while complex-units were dominated by STC filters. Finally, the ratio of the amount of information in the first asymmetric filter with the sum over the next twenty symmetric filters varied widely (mean 25, range 5-38). This is in line with our previous finding of a continuous and not bimodal distribution of simple- and complex-like units.

2.3.9 Hidden units: computing motion through recurrence

Finally, we investigated the feedforward connectivity from the input to the hidden units and the lateral connectivity among hidden units. We defined a hidden unit's direct weight as the pattern of weights connecting it to all input units. A hidden unit's indirect weight was defined as the average weight that connected the unit to the input units via the lateral connections of the hidden layer (see Materials and methods). Figure 12C shows the direct (solid) and indirect (dashed) input weights of two example simple-like hidden units, one tuned to the preferred direction (square) and one tuned to the anti-preferred direction (diamond) of the RMM output population. The first example hidden unit had a direct input that was Gabor-like with a peak at 5° and an indirect input that was also Gabor-like, but shifted to the left by 0.5°. For the second example hidden unit, the Gabor-like indirect input was shifted 0.1° to the right of the direct input. This shows that the network self-organized spatially asymmetric recurrent weights; such a connectivity pattern has been shown to generate direction selectivity (Mineiro & Zipser, 1998).

An alternative hypothesis of the mechanism underlying direction selectivity in the RMM starts from the observation that spatially shifted (dx) inputs through recurrent connections were always delayed by a single time step (dt; the time step of our simulations). This suggests that the hidden units could compute motion in the same way as proposed in the ME model: by the linear summation of two spatially shifted, temporally delayed inputs. Note that by considering only a single time step this view is in fact an inappropriate (non-recurrent) simplification of the recurrent network, even though it may seem a natural simplification within the ME framework. We follow this line of reasoning here only because its predictions are informative. This scheme predicts that the preferred speed of the units is given by dx/dt, where dx is the spatial shift between the direct and indirect input, and dt is the time delay between the direct and the indirect input which was 13 ms in our simulations. In Figure 12D we plot this linear summation prediction (dx/dt) against the actual preferred speed for all simple-like (black) and complex-like (gray) units whose spatial shifts could be estimated reliably (see Materials and methods).

The wide scatter of the data points clearly shows that the population as a whole does not follow the prediction based on the ME model. While some simple-like units are relatively close to the slope-1 line, most are not, and the linear summation scheme either overestimated or underestimated the real preferred speed. This mismatch is even more pronounced for most of the complex-like units where, surprisingly, the linear summation scheme often predicted the opposite direction of motion (grey data points in the second and fourth quadrant).

One way to phrase this result is that the recurrent network connectivity changes the effective dt from the fixed 13 ms delay generated by the simulation time step, and/or the effective dx from the fixed distance between the direct and indirect inputs. Consistent with the view that complex-like units are more driven by the recurrent network dynamics than the simple-like units, which are dominated by the afferent input, these dynamic changes in the spatiotemporal response properties are more pronounced in complex-like units than simple-like units.

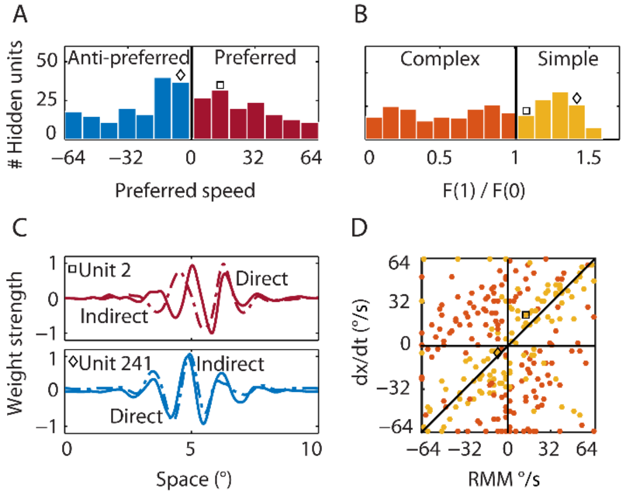

Figure 12. Properties of the hidden units. A) Preferred speed. Roughly half of the hidden units were tuned to the preferred (positive values on the x-axis) and half to the anti-preferred motion direction (negative values on the x-axis) of the output units. Whereas many units that were tuned to the preferred direction were tuned to the average preferred speed of the output units, many units that were tuned to the anti-preferred direction were tuned to lower than the preferred speed. The square indicates the preferred speed of the first example hidden unit and the diamond the preferred speed of the second example hidden unit in (B) and (C). B) Relative response modulation. We classified simple and

complex hidden units according to the conventional cutoff value of F1/F0 = 1. The square and diamond use the same convention as in (A). C) Direct and indirect inputs for two example hidden units. The first unit (top; square) was tuned to the preferred motion direction of the RMM output units. The direct (solid) and indirect (dashed) input weight distribution were Gabor-like and the indirect input was shifted to the left of the direct input. The second example unit (bottom; diamond) was tuned to the anti-preferred motion direction of the RMM output population and had an indirect peak input shifted to the right of the direct peak input. D) The preferred speed of the simple (black) and complex (gray) hidden units as predicted by a linear summation scheme (y-axis) plotted against the actual preferred speed (x-axis). Positive values show the preferred speed in the preferred direction. Negative values show the preferred speed in the anti-preferred direction. Most simple-like hidden units (like the two example units) were not on the diagonal. This shows that the RMM does not use a motion-energy like linear computation with fixed delays between spatially shifted detectors to compute velocity. This is even more pronounced for the complex-like units where the real preferred motion direction was often opposite to what the linear summation scheme predicts (opposite quadrants).

In the RMM all hidden units project to the output units. As a consequence, both simple-like and complex-like hidden units contributed equally to the velocity tuning of the output units. As this may appear to conflict with the evidence that MT cells receive mainly V1 complex input (Movshon & Newsome, 1996), we also developed a simple modification of the RMM with two layers of recurrently connected hidden units with feedforward connections between them. This network performed comparably to the RMM studied in detail here, but its first hidden layer developed mainly simple-like units, while the second developed mainly complex-like units. This shows that anatomical constraints can easily be incorporated in the RMM.

2.4 DISCUSSION

We showed that a recurrent network can generate the velocity-tuned response dynamics measured in area MT. This network used only a single delay, but nevertheless generated output units with a wide range of speed preferences. When the output units were tested with noise stimuli, they had slanted space-time filters with symmetric nonlinearities for the preferred and anti-preferred direction of motion, much like a feedforward ME network. The RMM, however, captured the full time course of velocity tuning, while the feedforward approximation could not. This strongly suggests that higher than second order spatiotemporal interactions play an important role in motion detection. The hidden units of the RMM showed a continuum of simple- to complex-like properties consistent with those found along the motion pathway of the primate brain. The velocity tuning of these units did not arise from the linear summation of spatially shifted and temporally delayed inputs (as in the ME model), but instead relied on asymmetric spatial connectivity and the nonlinear operations embedded in the recurrent interactions to become sensitive to a wide range of velocities.

After discussing some of the practical limitations of our modeling effort, we discuss the origin of delays in the RMM, the importance of considering the full time course of motion selective responses, and compare the RMM to the ME model.

2.4.1 Limitations

Our experiments used a stimulus with a diameter of 10° and a monitor refresh of 75Hz (13 ms). This naturally determines the spatial and temporal bounds on the motion tuning we could find (and then model). For instance, due to aliasing, stimuli moving at 64°/s on a 75Hz monitor generate limited directional motion signals, hence we did not attempt to model MT neurons with very fast speed preferences. Similarly, very slow movements are affected more by the discretization of space (the limited number of input neurons) and our representation of the random dot patterns removed high spatial frequency and therefore some low speed information. In other words, the particular choices we made to approximate the spatiotemporal properties of the stimuli used in the experiment (e.g. RF size, low-pass filters, simulation time step) limited the range of neurons that the RMM could feasibly model. Our selection of 26 neurons (for instance from the middle of the range of speed preferences) was partially based on that. Hence, we do not claim that the specific RMM used here can model the response of any MT neuron; neurons with very high preferred speed, for instance, would likely require an input layer spanning a larger part of space. Interestingly, this is consistent with the finding that preferred speeds increase with RF size (Orban, 1986) and the suggested early stage of speed tuning in the model of (Chey, Grossberg, & Mingolla, 1998).

Our fixed 13 ms simulation time step is a crude abstraction of the dynamics of the visual system. This window was mainly chosen for practical reasons. First, the random dot patterns were displaced every 13 ms; by choosing a simulation time step of (at most) 13 ms we could simulate the response to each pattern that was shown to the neuron. Shorter simulation time steps would have come at rapidly increasing computational cost, but also requires us to use spike count estimates from shorter windows, which would have made these estimates less reliable. Finally, we note that the 13 ms time step is within the approximate temporal integration range of 10-30 ms for pyramidal cells in cortex (Marmarelis & Marmarelis, 1978), hence it does not seem inappropriate to lump activity within such a window. We acknowledge, however, that interesting structure may be found in the time course of motion selective neurons at even shorter time scales.

2.4.2 Delay lines

The RMM provides a proof of principle that a network, in which all neurons have the same intrinsic delays, can nevertheless generate motion sensitivity with a wide range of preferred speeds. This speed tuning is the result of spatially asymmetric connections (input weights) and nonlinear recurrent dynamics (lateral and recurrent weights) that generate a range of effective delays (Mineiro & Zipser, 1998; Sabatini & Solari, 1999).

A simple linear combination of the spatially asymmetric (dx) and the temporally delayed (dt) inputs as envisaged in the feedforward ME model did not provide an accurate account of velocity tuning for the hidden units of the RMM (Figure 12D). This reinforces the point that the properties of a single unit in a network dominated by recurrent connections cannot be fully understood on the basis of a snapshot of its input. Moreover, the origin of the temporal delays in the RMM is conceptually different from the feedforward ME model where classes of slow and fast neurons generate speed tuning, or the Reichardt detector with its explicit delay lines. Importantly, the RMM's mechanism is not contradicted by the finding that the inputs to direction selective simple cells look as if they originate from fast and slow populations (De Valois & Cottaris, 1998). In the RMM the inputs look like that as well, but because all units have the same intrinsic delay (13 ms; our simulation time step) we know that an interpretation in terms of two populations with different intrinsic properties is incorrect. Hence at the very least our findings serve as a caveat to the interpretation of these empirical findings.

In other recurrent network based motion models (Maex & Orban, 1996; Suarez et al., 1995) temporal delays were implemented by the use of slow (NMDA) and fast (GABA) synaptic transmission. Given that these intrinsic parameters are fixed, this approach can only generate different speed preferences by changing the spatial offset between inputs. This is contradicted by the finding that both spatial and temporal offsets affect the computation of velocity (Koenderink, van Doorn, & van de Grind, 1985). Our findings (Figure 12) show that a recurrent network does not require a range of network delays to generate a range of effective delays (and hence speed preferences); it achieves this by the judicious choice of network connectivity strengths. This is in line with (Clifford et al., 1997; Clifford & Langley, 2000) who suggested that the temporal filter of the Reichardt detector and the ME model can be implemented recursively and that this significantly reduces computational and storage costs.

2.4.3 Time course

The time course of velocity tuned responses has received little attention in models, but can be quite revealing of the underlying mechanisms. Our analysis in Figure 10C-D, for instance, shows that the best feedforward LN-model, unlike the RMM, cannot capture the full velocity tuning properties of the MT cells (R^2=0.3 and 0.6 (SI), R^2=0.5 and 0.8 (DSI), respectively). The same is true for the implementation of the ME model by Simoncelli and Heeger (Simoncelli & Heeger, 1998) which considers only steady-state velocity tuning. Because both the LN approximation to the RMM and the ME model rely only on spatiotemporal correlations up to second order, this mismatch strongly suggests that MT neurons are also driven by higher-order spatiotemporal correlations. The RMM, on the other hand, is sensitive to higher order correlations and can therefore reproduce the full time course of the single MT cells much more faithfully. This provides a novel insight into the phenomenon of sequential recruitment: motion selectivity in MT (Mikami, 1992) and motion detection performance (McKee & Welch, 1985) improves nonlinearly with the number of successive steps in an apparent motion sequence. Our model suggests that this phenomenon relies critically on the recurrent network dynamics.

2.4.4 Comparison to the Motion Energy Model

When analyzed with white noise methods, the RMM revealed filters and nonlinearities that were at least superficially consistent with the ME model (i.e. slanted in space time, a quadratic nonlinearity, and a form of motion opponency). While this provides support for the model (because such filters have been found empirically), it also makes an important conceptual point about the interpretation of the empirical data. Notably, finding such ME-like filters in real neurons does not prove that the underlying architecture is at all similar to the feedforward ME model.

We also found a number of deviations between the RMM and the idealized ME model (Adelson & Bergen, 1985)in each case the RMM properties are supported by the empirical data. First, reverse correlation of the output units as well as that of the hidden units revealed many more than four space-time filters. This is supported by the large number of slanted space-time filters in V1 simple and complex cells (Rust et al., 2005). The RMM also deviates from the idealized ME model in its use of motion opponency. The filters of the output and hidden units, for instance, were not mirror opposites of each other (Figure 8B) as would be expected from pure motion opponency. Of course this asymmetry can also be seen in the speed tuning curves of the MT neurons Figure 7A, but in addition it is compatible with the DS V1 cells from (Rust et al., 2005) and our previous finding that motion opponency in MT involves a competition among multiple Fourier components, rather than a strict inhibition between opposite velocities (Krekelberg et al., 2005).

2.5 Conclusion

A recurrent network can compute a representation of velocity in much the same way as the ME model, but without the need for separate classes of fast and slow neurons or synapses. In contrast to the ME model, the recurrent network also matches the temporal dynamics of a population of single MT cells, and makes use of higher-order spatiotemporal correlations in the input. Because it relies on the pervasive recurrent connections of visual cortex, and given that it contains hidden units that are similar to other neurons in the motion processing pathway of the primate brain, we believe it is a biologically plausible model of motion detection.

Even though we focused on motion detection here, the training of artificial recurrent networks on recorded neuronal responses may also be a generally useful approach to investigate other domains of sensory processing and higher cognitive function that require the representation of sequences and time, which is thought to depend critically on recurrent network dynamics (Elman, 1990).

3 Recurrent network dynamics; a link between form and motion

3.1 Introduction

Form perception is often described as the detection of corners and junctions (Das & Gilbert, 1999), contours (Heydt & Peterhans, 1989; Ito & Komatsu, 2004; Lee & Nguyen, 2001) figure-ground segregation (Qiu & Heydt, 2005), arcs and circles (Hegdé & Van Essen, 2000). Each of these high-level concepts, however, can also be understood in terms of mathematically precise spatial correlations between three or more points (multipoint correlations). For instance, four-point correlations signal contours (Heydt & Peterhans, 1989; Ito & Komatsu, 2004; Lee & Nguyen, 2001), even illusory ones (Heydt, Peterhans, & Baumgartner, 1984), and three-point correlations provide information on figure/ground segregation (Victor & Conte, 1991; Yu et al., 2015). This suggests that a framework based on multipoint correlations can be fruitful to understand form perception.

The power of this framework is demonstrated by a cluster of recent findings. First, humans are highly sensitive to multipoint correlations that vary most in natural images (Hermundstad, Briguglio, Conte, & Victor, 2014; Victor & Conte, 1991). Because humans are most sensitive for patterns that are least predictable, thus carry most information (Hermundstad et al., 2014; Tkacik, Prentice, Victor, & Balasubramanian, 2010), this is evidence for a form of efficient coding (Barlow, 1961; Doi & Lewicki, 2014; Hateren, 1992). Second, while only some neurons in area V1 are selective for multipoint correlations, a significant fraction of V2 neurons respond selectively to visually salient three- and four-point correlations (Yu et al., 2015) Moreover, naturalistic textures – which are distinguished from their Gaussian-noise analogs on the basis of multipoint correlations – lead to distinctive responses in V2, both in the human and the macaque (Freeman et al., 2013). In this paper we propose a novel mechanism by which neurons in V1 and V2 generate such selectivity.

Previous approaches to understand form processing in early visual areas have relied on feedforward models that combine multiple linear filters through static nonlinearities (Chichilnisky, 2001; Rust et al., 2004; Simoncelli et al., 2004). In principle such models can be selective for capturing multipoint correlations. However, as we show below, for the specific dataset we aimed to model, this approach did not fare well. This may in part be due to the poor match between the single-stage feedforward processing in LN models and the multi-stage processing and abundance of recurrent connections in the visual system. We, therefore, developed an alternative approach based on a four-layer artificial neural network with locally recurrent connectivity. This artificial neural network faithfully reproduced neurons' selectivity for multipoint correlations and generalized beyond the V2 data set that was used to fit the model.

New insights into early visual form processing resulted from a detailed investigation of the fitted artificial neural network. The model self-organized network elements with response properties surprisingly similar to individual neurons recorded in V1 and V2, including selectivity for visually salient three- and four-point correlations and characteristic response dynamics. Finally, we tested the tuning properties of the network units to dynamic stimuli and found that many neurons were tuned for motion and that four-point selectivity was strongly correlated with selectivity for motion. This leads to the novel and testable prediction that complex form analysis and motion tuning are closely intertwined at the single neuron level as early as V1 and V2.

3.2 MATERIALS AND METHODS

3.2.1 Experimental Data

The experimental data were obtained using tetrode recordings in areas V1 and V2 of 14 anesthetized and paralyzed macaques. All procedures were approved by the Weill Cornell Medical College Animal Care and Use Committee and were in agreement with the National Institutes of Health guidelines for the humane care and use of laboratory animals.

We recorded 269 neurons in V1 and 153 neurons in V2 and confirmed the recording sites using electrolytic lesions at the conclusion of the experiment. In V1 we classified 32 cells as supragranular, 153 cells as granular and 71 cells as infragranular. In V2 we classified 32 cells as supragranular, 34 cells as granular and 57 cells as infragranular. This dataset consisted of all of the recordings reported in (Yu et al., 2015), except for the V1 (13/269) and V2 (30/153) neurons for which laminar identification was uncertain. Details concerning animal preparation, electrophysiological procedures, stimulus alignment, spike-sorting, response analysis, and histology are provided in (Yu et al., 2015).

3.2.2 Visual Stimuli

All stimuli were checkerboards, consisting of a 16x16 array of black and white checks. Checkerboards were either random (check colors assigned independently and with equal probability to black or white), or constructed to contain only spatial correlations of a specific spatial configuration and order (Figure 13). The latter are designated MSCT's (multipoint spatial correlation textures). We studied 6 MSCT classes: two classes contained visually salient three-point correlations (*white triangle* and *black triangle*), two classes contained visually salient four-point correlations (*even* and *odd*), and two classes contained four-point correlations that are not visually salient (*wye* and *foot*). Stimuli from these six classes were generated via a Markov recurrence rule (Victor & Conte, 1991, 2012). We presented 1024 examples (two repeats each) per MSCT class for 320 ms, interleaved in a pseudorandom sequence. It is important to note that for each MSCT, the specific multipoint correlations are fixed, and there are (on average) no correlations of lower orders (e.g. the *even* stimulus class has a specific fourth-order correlation, but does not have first- (mean luminance), second- (power spectra/spatial frequency content) or third-order correlations). Put differently, these classes form a basis to study the influence of each kind of multipoint correlation.

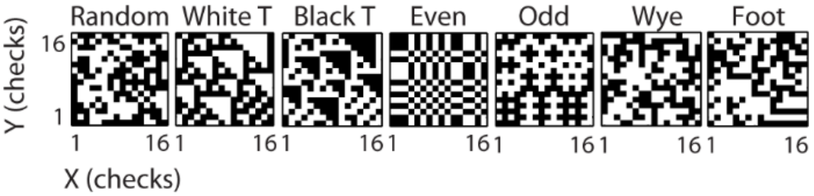

Figure 13. Multipoint spatial correlation (MSCT) stimuli. One example texture is shown for each of the texture classes. The visually salient white triangle (White T) and black triangle (Black T) textures differ from the random textures in their three-point correlations. The visually salient Even and Odd textures and non-visually salient Wye and Foot textures differ from the Random textures in their four-point correlations.

3.2.3 Data analysis

3.2.3.1 Linear-nonlinear model

In the linear-nonlinear (LN) model we adapted from (Chichilnisky, 2001; Rust et al., 2004; Simoncelli et al., 2004) the visual input is first linearly filtered by one or more filters, each filter output is transformed by a static nonlinearity, and these outputs are then summed. We used the spike triggered average (STA) and the spike triggered covariance (STC) methods to estimate the filters (Chichilnisky, 2001; Rust et al., 2004; Simoncelli et al., 2004) using the full set of stimuli (1024 examples, 7 classes, 2 repeats) and the mean response over 40-200 ms after stimulus onset. Based on the STA and STC we then estimated the information captured by the maximally informative filters using the iSTAC method (Simoncelli et al., 2004). For display purposes (Figure 16), these linear filters were low pass filtered with a 2-dimensional Gaussian (σ = 2 input stimulus checks). Finally, we determined the nonlinearity associated with each filter by dividing the histogram of the projected spike triggered ensemble by the histogram of the projected raw stimulus ensemble, over four standard deviations away from the mean. This procedure assumes separability of the filter dimensions (Simoncelli et al., 2004).

We estimated the performance of each LN model separately on the 1024 examples per MSCT class that were used to estimate the LN models (train set) and 10,000 newly generated examples for each MSCT class (test set). For each MSCT stimulus we calculated the model output and averaged the response over all textures in an MSCT class (separately for train and test sets) to obtain an MSCT tuning curve. Model performance was defined as the Spearman correlation between the model tuning curve for the train and test sets, and the experimentally measured tuning curve (based on the train set).

3.2.3.2 Recurrent form analysis model

3.2.3.2.1 Elman recurrent neural network

The two-stage recurrent form analysis model (RFAM) was based on the Elman recurrent neural network (Elman, 1990) implemented in the MATLAB Neural Network Toolbox (version 4). Units in such an artificial network are considered a crude approximation of a neuron or a group of neurons (Figure 14). The units were interconnected with adjustable weights simulating synaptic connections with variable strength. Each unit also had an adjustable bias value. The network had one input, two hidden, and one output layer. The input layer consisted of 256 units that each simulated one of the 16 by 16 checks of the experimental stimuli. The input layer was fully connected to the first hidden layer in a feedforward manner. The first hidden layer had 100 units that were fully connected to the 100 units of the second hidden layer, which were fully connected to the output layer of the RFAM, both in a strictly feedforward manner. In addition, the hidden units of both layers were laterally/recurrently connected to all hidden units within their layer. The output for each unit (i) was calculated by first determining the weighted sum of its inputs plus the bias value: $\mathrm{x_i} = \sum_{k}^{i} \mathrm{w_{ik}y_k} + \mathrm{b_i}$, where the index k runs over all units that are connected to unit i, and then passing this through a sigmoid transfer function: $y_i = \ 1/(1 + e^{-\mathrm{x_i}})$.

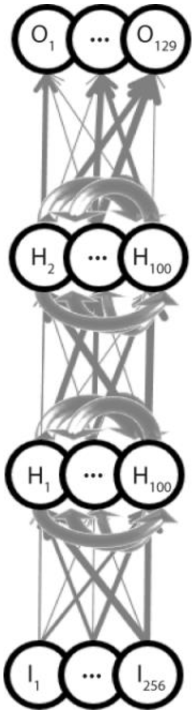

Figure 14. Recurrent Form Analysis Model (RFAM). The two -stage recurrent neural network had 256 input units (one for each check in the stimuli shown in Figure 13) that were all-to-all connected to the units of the first hidden layer. They, in turn, were all-to-all connected to the units of the second hidden layer. Both hidden layers had 100 units which were recurrently connected to all units within the same layer (thick gray lines). The units of the second hidden layer were all-to-all connected to the output unit(s). The weights of the connections between the neurons were adjusted in an iterative procedure (backpropagation through time) to simultaneously reproduce the output of each of the 123 V2 neurons in the 123 output units.

We developed two recurrent models. The first (RFAM) was trained to capture the response of all 123 V2 neurons (irrespective of their laminar location) in an output layer with 123 units. In the model network these output units are not connected; their interaction arises only from sharing a common set of hidden units. Figure 15 shows six examples of single neuron responses that these output units were trained to capture.

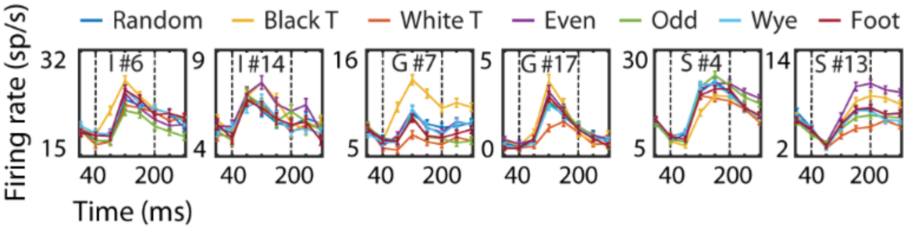

Figure 15. V2 example cells. The response of six V2 example cells (I = infragranular, G = granular, S = supragranular) to the MSCT textures over time (mean over 1024 examples and two repeats, error bars reflect the standard error). Dotted vertical lines indicate the time window used for further analyses and modeling. This figure shows that V2 MSCT selectivity was diverse and that the time course was complex.

The second RFAM model was trained to capture the average response of the supragranular V2 neurons (the neurons with strongest multipoint tuning). We refer to this model as the RFAM population average; RFAMpa. RFAMpa had a single output unit; its activity was trained to reproduce the average activity of all V2 supragranular neurons (V2pa).

3.2.3.2.2 Output patterns

The RFAM was trained to reproduce V2 responses, in the language of artificial neural networks these are called the target patterns, or, because they are the responses of the output units, output patterns. We chose to train the network on what we consider to be the most interesting phase of the response; the time period when selectivity for MSCT arises in most V2 neurons (40-200 ms; see Figure 15, marked by the dotted lines). This period excludes the initial descending response of approximately 40 ms that was most likely due to the previous stimulus that was presented as part of a stimulus stream without blank intervals. And, it also excludes the response changes that happen on a slower time scale, presumably due to adaptation processes.

Within the time period of interest, we binned the spiking response in 40 ms time bins to create output pattern sequences of length five. We normalized these responses to a suitable range for the artificial neural network (between zero and one) by first subtracting the minimum firing rate and then dividing by the maximum firing rate over all time bins, conditions, and neurons.

3.2.3.2.3 Input patterns

The input patterns presented to both RFAM designs matched the set of 1024 examples per class used for the electrophysiological experiment. The binary 16x16 stimuli were spatially low pass filtered with a 2-dimensional Gaussian ($\sigma = 2$ checks) to generate a continuous representation, and to approximate the likely input to cortical neurons, which would be filtered by the lens, retina, LGN, and other sources of blur. Although this low pass filtering introduces second-order spatial correlations in the textures, this is equal for all MSCT classes and does not affect multipoint correlations. Just as in the experiments, the same, static pattern was presented for each of the five 40 ms time bins of a simulated trial. Between trials the activity in the network was reset to zero to avoid spurious interactions between successively presented training patterns.

3.2.3.2.4 Training phase

Before training the network, we initialized the weights and bias values of all layers using the method of (LeCun, Bottou, Bengio, & Haffner, 1998). In the training phase, we randomly chose one of the input patterns and presented this to the network and calculated the response of all units in the network for five time steps. Next, we calculated the error as the mismatch between the response of the 123 output units and the 123 V2 cell responses. (For RFAMpa, the error was defined analogously as the mismatch between the single output unit and the V2pa). This error was then used to modify the connection weights in the network using error back-propagation-through-time. This process was repeated five million times (epochs). Monitoring the error over time showed that this training period was sufficient to reach convergence (i.e. further training contributed little to a reduction in error, see Figure 18A). Network parameters were then frozen and we investigated the trained network.

As is the case with all artificial neural networks, design choices such as the number of layers, neurons per layer, and number of training epochs proceeded largely by trial and error. For instance, we discarded networks with smaller numbers of hidden units for which the training algorithm failed to converge to a solution. The findings reported here, however, were robust to changes in these choices, and were found reliably in all networks trained on these data (even though each training procedure started from different random initializations of network connectivity).

3.2.3.3 Texture tuning index

We quantified selectivity for MSCT with the mean response over 1024 experimental examples per MSCT class and time (and two repeats for the V1 and V2 cells). We calculated the texture tuning index (TTI) for each of the six MSCT classes as the absolute value of the Michelson contrast between the mean response to the class (x) and the mean response to the *random* textures: $\text{TTI}_x = |(x - \textit{random}) / (x + \textit{random})|$. A TTI of zero corresponds to no selectivity, higher TTIs represent increasing selectivity.

3.2.3.4 Temporal dynamics

We used principal component analyses (PCA) to investigate the time course of the neural and model unit responses. First, we calculated the mean response over the 1024 experimental examples and the MSCT classes. To align the V1 and V2 responses with the hidden units (which do not have an afferent delay), the former were shifted by their average onset delay (40 ms). All time courses for each neuron and unit were then normalized with a division by the maximum response over time. Next, all hidden units and all V1 and V2 neurons were collected in a single matrix and PCA was applied to this matrix. We quantified hidden units' and V1 and V2 neurons' temporal dynamics by their projection (Figure 21B-C) onto the first two principal components (Figure 21A).

3.2.3.5 Texture, orientation and motion tuning

To compare response properties for dynamic stimuli with response properties for MSCT classes, we defined a texture response (TR) that includes positive and negative values. For each hidden unit, TR_x was the response in the first five time steps (200 ms) averaged over 10,000 textures of one of the MSCT classes (x) minus the mean response averaged over 10,000 *random* textures. A positive or negative TR_x, therefore, reflects an increased or decreased activation for MSCT class-x compared to the *random* class, respectively.

To estimate the orientation tuning properties of the hidden units, we created oriented stimuli using 1-dimensional binary random noise values (16 values) replicated in the y-direction (16 values), and rotated these 2-dimensional patterns with one of 18 angles between 0° and 180°. Just as the MSCT patterns these were low pass filtered with a 2-dimensional Gaussian ($\sigma = 2$ checks). The same pattern was presented to the model for five time steps (200 ms). We define the orientation response (OR) for each hidden unit based on the mean response over 10,000 trials and time. The OR was the maximum difference between the 18 orientation conditions and the mean over all (baseline). Similar to the TR, negative or positive OR values reflect a reduced or increased activation for the preferred orientation compared to the overall mean over all orientation conditions.

We estimated the motion tuning properties for each hidden unit with 10,000 low pass filtered 2-dimensional *random* textures described previously. For translational motion, we moved the noise patterns with one of seven speeds (0, 0.5, 1, 2, 4, 8, 16 checks/40 ms) in one of four directions (*upward*, *rightward*, *downward* or *leftward*) over five time steps (200 ms). The V2 cells had different receptive field (RF) sizes and the check sizes were scaled to fit within each neuron's RF. We can, therefore, not assign a specific speed in °/s to our simulations. However, for the typical RF size of around 1° and a time step of 40 ms our simulated speeds are equal to 0 to 25°/s.

For rotational motion, we rotated the noise patterns with one of nine speeds (0, 0.5, 1, 2, 4, 8, 16, 32, 64, 128 °/40 ms) in one of two directions (*clockwise* or *anti-clockwise*) over five time steps (200 ms). Tuning curves were generated by taking the mean response over trials and time (e.g. example units in Figure 22). We define the translational speed tuned response (SRt) and rotational speed tuned response (SRr) as the maximum difference between the responses to all motion conditions compared to response to the stationary condition (e.g. speed 0 °/40 ms, see Figure 22B-C). Analogous to TR and OR, negative or positive SR reflect reduced or increased activation for dynamic stimuli compared to stationary stimuli. Finally, we define the direction (DRt) and rotation direction response (DRr) as the maximum difference between the response to the four direction conditions at the preferred or anti-preferred (rotation) speed.

3.3 RESULTS

After giving an overview of the experimental data, we first present an attempt to capture the computational principles underlying tuning for multipoint correlations using an established method based on feedforward processing. This method has previously been used successfully to reveal unexpected complexity in orientation selective V1 cells (Rust et al., 2005). As we will show below, however, that approach fails to explain the data at hand. This motivated us to develop a novel approach using a recurrent network model, which is the focus of the third and major part of this section.

3.3.1 Experimental data

We recorded from 269 neurons in anesthetized, paralyzed macaque V1 and 153 neurons in V2. Based on histological verification we classified 32 V1 cells as supragranular, 153 cells as granular, and 71 cells as infragranular. In V2 we classified 32 cells as supragranular, 34 cells as granular, and 57 cells as infragranular (Yu et al., 2015). This dataset consisted of all of the recordings reported in (Yu et al., 2015), except for the V1 (13/269) and V2 (30/153)] neurons for which laminar identification was uncertain. We stimulated the cells with example textures of seven 2-dimensional texture classes that isolate multipoint correlations previously studied psychophysically (Hermundstad et al., 2014; Tkacik et al., 2010; Victor & Conte, 1991, 2012). Examples of the multipoint spatial correlation textures (MSCT) are illustrated in Figure 13. For the *random* textures, check colors were assigned white or black independently. The *white triangle* and *black triangle* textures isolate the extremes of the visually salient three-point correlations and the *even* and *odd* textures isolate the opposite extremes of the visually salient four-point correlations (Hermundstad et al., 2014). Finally, the *wye* and the *foot* textures have four-point correlations that are not visually salient (Victor & Conte, 1991).

Yu et al (Yu et al., 2015) reported that some V1 and many V2 cells showed selectivity for MSCT. To obtain a robust measure of this selectivity that could be compared with our models, we determined an MSCT tuning curve (the average response to the example stimuli from an MSCT class), separately for two randomly chosen halves of the data (512 examples per class each). For each randomly chosen 50/50 split we calculated the correlation between the two tuning curves and repeated this process 5000 times (drawing new random 50/50 subsets each time). Throughout this paper we define consistency as the mean of the distribution of correlations over these 5000 sets. A neuron with multipoint tuning that generalized to all examples of the MSCT classes would have a consistency of 1. In V1, the consistency quartile range [25th percentile, 75th percentile] was [0.06, 0.42], in all of V2 it was [0.33, 0.60], and in supragranular V2 it was [0.46, 0.70]. This shows that a substantial fraction of neurons, and especially those in the supragranular layers of V2, have robust tuning for multipoint correlations. For other measures of tuning and a detailed analysis of the robustness of multipoint tuning in V2 neurons, we refer to (Yu et al., 2015).

Our goal here was to uncover computational principles that could underlie the tuning for multipoint correlations observed primarily in V2.

3.3.1.1 Linear-nonlinear model

In an LN model each subunit receives the same input, which is passed through a (linear) filter and then through a static nonlinearity. Here we used the information theoretic spike triggered average and covariance analysis (iSTAC) method (Pillow & Simoncelli, 2006) to estimate the most informative subunit filters as well as their nonlinearities (Materials and Methods). The input to the iSTAC method was the collection of 1024 example textures for each of the seven MSCT classes and the output was the mean firing rate (FR) evoked by the neuron in a time window 40-200 ms after stimulus onset. We note that the iSTAC procedure does not attempt to determine the dynamics of the initial linear filters and that there is a second linear stage that simply sums across each LN component to generate a single output based on multiple, parallel LN pathways. Because this second stage has no additional free parameters, we refer to the model as an LN model (even though it could technically be considered an LNL model).

Figure 16A shows the STA and the eleven most informative filters in the space spanned by the STA and spike triggered covariance (STC) (Pillow & Simoncelli, 2006) for one of the V2 supragranular cells (#13). The STA shows that this cell has a polarity-sensitive patch in the center of the cell’s RF. The similarity between the STA and filter #1 (Figure 16A) shows that most of the information was carried by the STA. The next two most informative filters (#2 and #3, Figure 16A) had orthogonal orientation sensitivity. In contrast to filter #1, the orientation sensitivity was polarity insensitive; both stimuli that matched a filter or that matched the polarity inverse of a filter evoked increased responses compared to the mean FR. The next eight filters were excitatory (#5, #8 and #9; textures that match these filters increased the FR above the mean) or suppressive (#4, #6, #7, #10 and #11; textures that match these filters decreased the FR below the mean), but none had obvious spatial structure.

In the feedforward view of visual processing, this set of filters and their corresponding nonlinearities, accounts for the output of a V2 cell. The question we asked is whether this model can explain sensitivity for complex form. Figure 16B shows the neuron's MSCT tuning curve (mean response over 1024 examples, two repeats, and time window 40-200 ms after stimulus onset). The neuron was selective for MSCT (ANOVA, p<0.0001) and specifically for the visually salient three-point MSCT with a decreased response for *white triangle* textures (post hoc t test, p<0.0001) and an increased response for *black triangle* textures compared to the *random* textures (post hoc t test p<0.008). This V2 cell also responded strongly to the visually salient four-point MSCT with an increased response for the *even* textures compared to the *random* textures (post hoc t test p<0.0001). The consistency of MSCT tuning (see above) for this neuron was very high: r=0.93, showing that its tuning generalized almost perfectly to all examples of the MSCT classes.

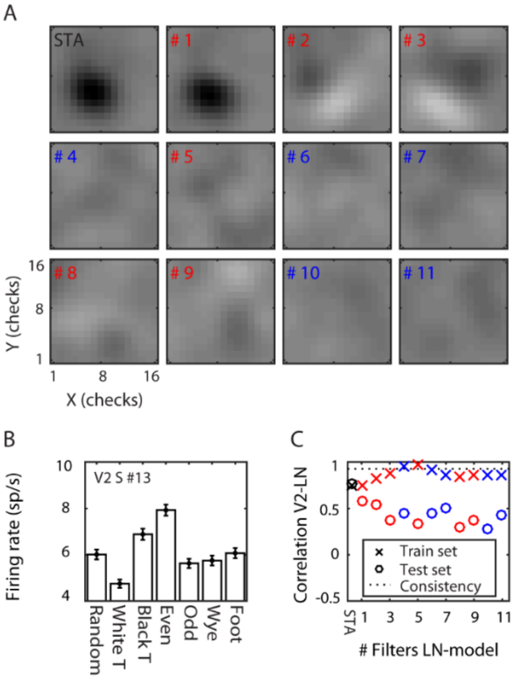

Figure 16. LN model of a V2 supragranular example cell (#13). A) Linear filters. The spike triggered average (STA) and the 11 linear filters ordered by the amount of information they carry (filter numbers show rank order). Red/blue indicates filters that increase/decrease firing rate above/below the mean of the cell. B) MSCT selectivity. Mean response to the seven MSCT classes. The error bars indicate standard error over examples. This supragranular V2 cell responded selectively to three-point and four-point textures. C) Performance of the LN model. Correlation between the neuron's texture tuning curve and the tuning curve of the LN models with increasing number of filters (x-axis). Performance is shown separately for the train set (crosses) and for generalization to a test set (open circles). The dotted line shows the tuning consistency of the V2 neuron (see main text for details). This figure shows that the LN model captured the tuning in the train set but not the ability of the V2 neuron to generalize across examples of the MSCT classes.

We quantified the LN model's ability to reproduce MSCT tuning as the Pearson correlation between its MSCT tuning curve and the MSCT tuning curve of the corresponding neuron. This performance was first calculated based on the response to the 1024 examples per MSCT class that were also used to estimate the LN model (train set; training tuning curve; training performance). To assess the model's ability to generate consistent MSCT tuning for stimulus examples that were not part of the training set, we also generated a model MSCT tuning curve (generalization tuning curve) based on the simulated response to 10,000 new examples per MSCT class (test set). The correlation between the generalization tuning curve and the neural tuning curve defines the generalization performance. To assess the contribution of each of the filters, we calculated model performance separately for models that included only the STA, only the first (most-informative) filter, only the first two most informative filters, up to the first fifteen most informative filters.

Figure 16C shows the performance of the LN models for the train set (crosses; training performance) and the test set (open circles; generalization performance). For this neuron, the STA model captured a considerable amount of the MSCT selectivity (r=0.76), and a five-filter LN model resulted in almost perfect training performance (r=0.98). However, the model fared poorly on new example textures, with generalization performance around r=0.5 regardless the number of filters in the model.

For comparison, the dashed line in panel C shows the generalization performance of the example neuron (r=0.93); clearly the LN model performed much worse than the example neuron, suggesting that many of the filters and corresponding nonlinearities did not capture the underlying regularity of texture tuning.

We estimated analogous LN models for each of the V1 and V2 cells. These models often explained a large fraction of the measured V1 and V2 MSCT selectivity in the training set (V1 mean r= 0.56 +/- 0.3 stdev; V2 mean = 0.57 +/- 0.28 stdev), but they did not generalize to new stimulus patterns drawn from the MSCT classes (even for cells that had highly consistent MSCT tuning). Figure 17 documents this for the V2 population, separately for models that included only the first 4 filters (which on-average had the best generalization performance) and models that included 15 filters. The lack of out-of-sample generalization implies that these LN models provide little insight into the computations underlying sensitivity to multipoint correlations and demonstrates the need for a different approach. We chose to pursue a recurrent network model (see Discussion).

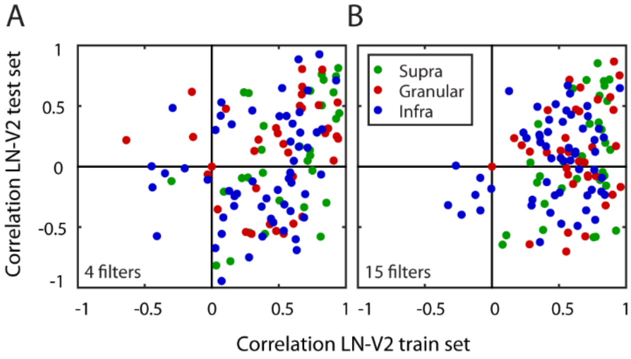

Figure 17. Performance of the LN models. Performance measures the correlation between MSCT tuning of each V2 neuron and its LN Model. Performance was calculated separately for a set of novel textures (test set) and plotted against the experimental stimulus set that

was used to estimate the LN model (train set). Each dot represents a single V2 neuron. Colors represent laminar origin of the neurons (see legend). A) LN models based on the first 4 most-informative filters. The selection of 4 filters was based on the population average performance on the test set; this was best for 4 filters, suggesting that additional filters mainly captured noise. B) LN models based on the first 15 most-informative filters; these models had good performance on the train set but generalized poorly. This figure shows that the LN models captured a significant fraction of the variance on the training set, but even when restricted to the best set of filters (A) generally failed to generalize to novel textures.

3.3.2 Recurrent form analysis model

Previously, we have shown that recurrent connections are able to capture higher-order space-time correlations and model motion tuning in neurons of the middle temporal area (Joukes et al., 2014). This, together with the fact that recurrent connections are ubiquitous in cortex led us to the hypothesis that a recurrent network could also be a basis for complex form analysis. We investigated this hypothesis with a recurrent neural network consisting of (artificial) neurons, all with identical intrinsic properties but modifiable feedforward and recurrent/lateral synaptic strengths (Elman, 1990). The recurrent neural network had 256 input units; one per check in the MSCT textures (Figure 13). The input units were connected in a feedforward manner to a first hidden layer (H1, 100 units) and the units in H1 were feedforward connected to the second hidden layer (H2, 100 units). H2 units, in turn connected feedforward to each of the 123 output units. Recurrent connections were introduced within H1 (each H1 neuron connected to all other H1 neurons) and, analogously, within H2.

In an iterative procedure, we presented one of the 1024 experimental example textures per MSCT class, simulated the response of the output units, calculated the mismatch between the recorded neural response of the 123 V2 neurons and the observed simulated response, and used this as the error signal in the back-propagation-through-time algorithm to adjust the weights in the network (Materials and Methods). We refer to this model as the recurrent form analysis model (RFAM). The first step was to reproduce the MSCT selectivity and temporal dynamics of the recorded cell responses.

We quantified performance of the RFAM output units separately for the train set (1024 experimental examples per MSCT class) and for a generalization set (10,000 new examples per MSCT class), just as we did for the LN model. Figure 18A shows how the performance (averaged over all 123 output units) improved with training. After five million training epochs, RFAM captured the MSCT tuning for textures in the train set (solid lines, r = 0.88) as well as textures that were not used to fit the model (dotted lines, r= 0.81). Figure 18B shows the performance of each of the RFAM output units on the train set plotted against the performance on novel textures that were not used to train the model (test set). This figure shows that the RFAM output units captured the essence of MSCT tuning observed in individual V2 neurons. In contrast to the LN model, generalization to the out of-sample test set was only slightly worse than the performance on the training set.

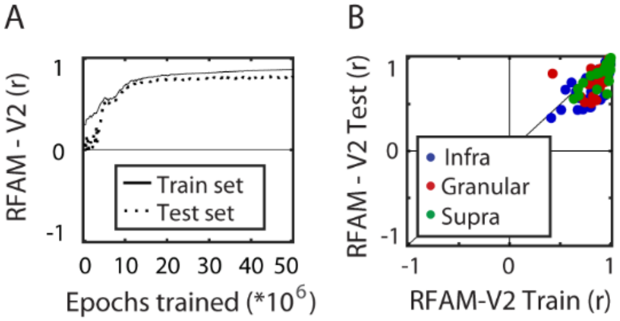

Figure 18. RFAM training and performance. Performance measures the correlation between MSCT tuning of the RFAM output units and the tuning of the target V2 neurons. Performance was calculated separately for the experimental stimulus set that was used to train RFAM (train set) and a set of novel textures (test set). A) Average performance of the 123 RFAM output units. High performance was reached for both the train set (solid line) and the test set (dashed line). B) Comparison of the performance of RFAM on the train set and the test set. Each dot represents a single V2 neuron. Colors represent laminar origin of the neurons (see legend). This figure shows that RFAM captured texture tuning in V2 neurons, and generalized to new examples from the texture classes.

Taken together these results show that the RFAM generalized to a large fraction of new examples of the MSCT classes, just as the V2 neurons. This suggests that two layers of recurrently connected neurons are sufficient to generate the tuning for multipoint correlations observed in V2.

3.3.2.1 Population average RFAM

In the analysis so far, we used the responses of each V2 neuron to train the RFAM and the LN models, including cells that had weak MSCT selectivity or low consistency over examples of an MSCT class. This allowed for a direct assessment of the models' ability to capture all experimental data. However, our main interest is not the specific observed texture tuning based on a subset of examples for any given MSCT class, but rather the underlying tuning rule for the full MSCT class. Combined with the goal to stay close to the experimental data, we chose to approximate this ideal with the average response of the 32 V2 supragranular cells (Figure 19A). We refer to this population average as V2pa. The V2pa had a high consistency (r=0.89; see Materials and Methods) indicating robust and consistent selectivity for all examples drawn from the MSCT classes. We modeled the V2pa with a single RFAM output unit (RFAMpa). For each of the 1024 example textures used in the experiment the target output used in the learning rule was the mean response of the V2pa across all 1024 textures of the same class used in the experiment. Put differently, V2pa and its model RFAMpa embody the consistent MSCT selective response observed on average in the supragranular layer of V2.

After training, RFAMpa had a strong preference for the *even* texture class and it responded with a transient-sustained response, just as the V2pa (Figure 19A). Most importantly, texture tuning of the RFAMpa network generalized well to textures not used in the training process (Figure 19B, train set r=0.98, generalization r=0.88). This shows that the RFAMpa solves the same computational problem that the supragranular V2 population solves; it consistently detects multipoint correlations in static images. Our next goal is to investigate how the model computes, and use this to generate a hypothesis and experimentally testable predictions for the analogous computations in the brain.

To answer the question how the recurrent network computes we analyzed the response properties of the hidden units. By focusing on a network that has been trained to produce a single output, we know that (by construction) the goal of each hidden unit's response is to bring the output unit closer to its target. This greatly simplifies the interpretation of hidden unit response properties and is a major advantage over analyzing the hidden units of the full RFAM network with 123 output units, in which each output unit has a slightly different target, and all hidden units contribute to each of those computations to some extent. Nevertheless, at the end of the results section we will return to the full RFAM network and show that the salient properties of its hidden units match those of the RFAMpa model.

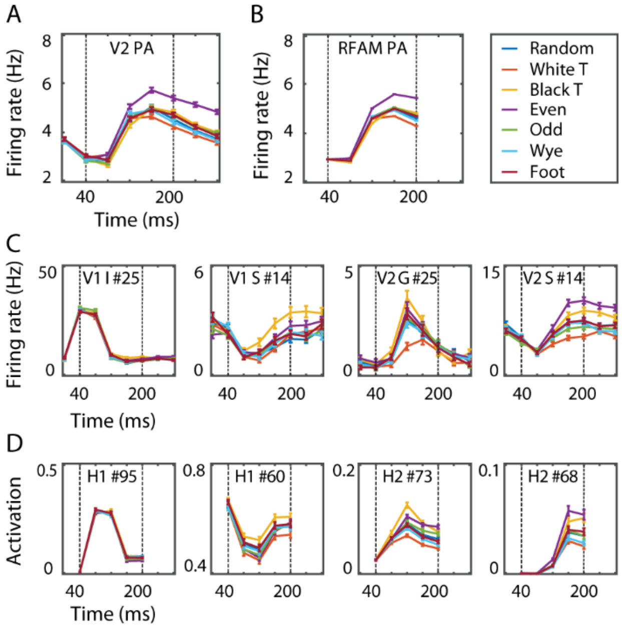

Figure 19. A) Dynamics of the V2pa response, averaged over the 1024 examples per MSCT class. B) Dynamics of the RFAMpa response, averaged over the same examples. C) Dynamics of two V1 and two V2 neurons. D) Dynamics of two H1 and two H2 RFAMpa units. Error bars indicate standard error over examples. This figure shows that, even though RFAMpa was trained only on static MSCT and only to reproduce the V2 population average responses (A), the hidden units self-organized diverse and complex MSCT selectivity and time courses similar to those observed in V1 and V2 isolated single cells.

3.3.2.2 Hidden units: texture tuning properties

We define a texture tuning index (TTI) as the relative change in average response to one of the MSCT classes compared to the *random* class (Materials and Methods). Figure 20 shows the TTI for each MSCT class, averaged over V1 neurons (panel B) and V2 neurons (panel D). This analysis confirms (using a slightly different metric) the results of (Yu et al., 2015); textures with visually salient high-order structure lead to responses distinct from those evoked by *random* textures, particularly in the supragranular layers of area V2. Panel A and C show the equivalent TTI for the hidden units of the RFAMpa. In the first hidden layer (H1), TTI’s were modest (panel A), but the second hidden layer (H2, panel C) had substantial texture tuning, in particular for those textures that are visually salient (*white triangle*, *black triangle, even,* and *odd*). There was little selectivity for the visually non-salient four-point textures (*wye*, *foot*) in either H1 or H2.

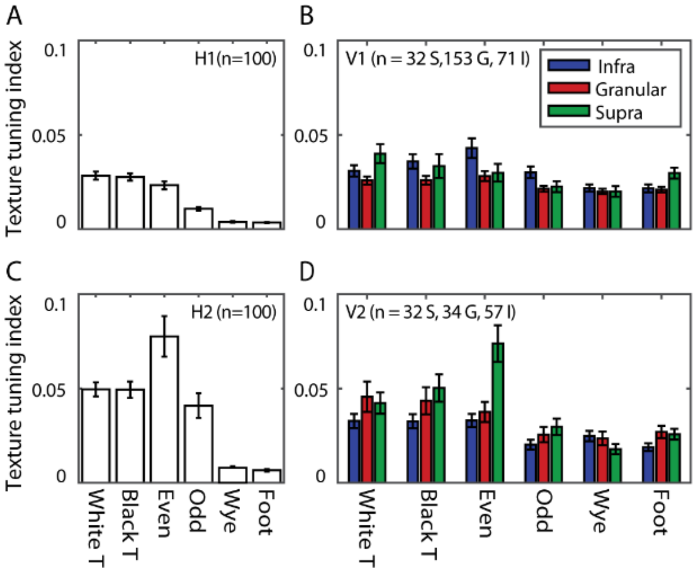

Figure 20. Texture tuning in the RFAMpa hidden layers (H1, H2) and V1, V2. Texture tuning quantified the average difference in response to examples from a given MSCT class and the random class. A) Mean texture tuning index for the six MSCT classes for the units in H1 of the RFAMpa. C) Same as A, now for H2. B) Mean texture tuning for the V1 neurons, grouped by cortical layer (I=infragranular, G=granular, S=supragranular). D) Same as B, now for V2. Error bars indicate standard error over units/neurons. Texture tuning was particularly strong for the visually salient MSCT (white triangle, black triangle, even, and odd) and stronger for H2 than H1. This qualitatively matches the MSCT selectivity properties of V1 and V2 neurons.

Taken together, this analysis shows that, although the only task given to the RFAMpa output unit was to reproduce the response of the V2pa (Figure 19A) at the population level, the training algorithm produced a network with hidden units whose MSCT selectivity was similar to that observed in V1 and V2 neurons. Note that the full range of V2 tuning properties is produced by the H2 units, even though the V2pa response was primarily selective for the *even* texture class, and showed little if any tuning for the other classes (see Discussion).

Thus far, we only analyzed the time-averaged responses. The V1 and V2 cells, however, had characteristic transient and/or sustained response properties (Figure 15). One of the main advantages of a recurrent network is that it can capture such dynamics more naturally than a feedforward model, and indeed, the RFAM and RFAMpa models were trained to reproduce the full time course of the response, not just the mean. This led us to investigate whether these dynamics play a role in generating selectivity for static stimuli with multipoint correlations.

3.3.2.3 Hidden units: temporal dynamics

Figure 19D shows the time course and MSCT selectivity of sample units in H1 and H2 and example neurons in V1 and V2 with similar tuning and response dynamics. Across the population of H1 and H2 units we observed many texture preferences and response dynamics. Notably, these preferences or dynamics could be quite different from those of the output unit (and V2pa). For instance, H1 unit #95 had almost no MSCT selectivity but a transient time course. H1 unit #60 and H2 unit #73 responded most strongly to *black triangle* textures; neither of these properties match the V2pa or RFAMpa. Similar properties, however, were observed in the individual V1 and V2 neurons. For instance, Figure 19C shows two V1 neurons (first two panels) and two V2 neurons (last two panels) with response properties that are qualitatively similar. These examples were hand-picked, but the following formal analysis confirmed a high degree of similarity between the dynamics of H1 & H2 on the one hand and V1 & V2 on the other.

We used principal component analysis (PCA) on the dynamics of all units (H1, H2) and all neurons (V1, V2) to extract a common basis for a low-dimensional description of the dynamics (Materials and Methods). Two components explained 84% of the variance in the temporal dynamics (Figure 19A), showing that little information is lost when describing each neuron by two numbers (the projections onto these two components). Figure 21B-D displays each of the subpopulations in this coordinate system and allows for a visual comparison and qualitative interpretation. First, H1 and H2 clusters clearly overlap with the V1 and V2 clusters, showing that their dynamics were generally similar. More quantitatively, 84% of the convex hull of V1 (Figure 21C) and V2 (Figure 21D) overlapped with the convex hull of H1 and H2 (Figure 21B). Apart from the general similarity between the hidden units and the neurons, this analysis also suggests a modest degree of hierarchical organization in the RFAMpa that appears to match that found in the brain. For instance, compared to the population of V1 cells, V2 cells had more positive projections on PC1 and PC2 and this is also seen in H2 compared to H1.

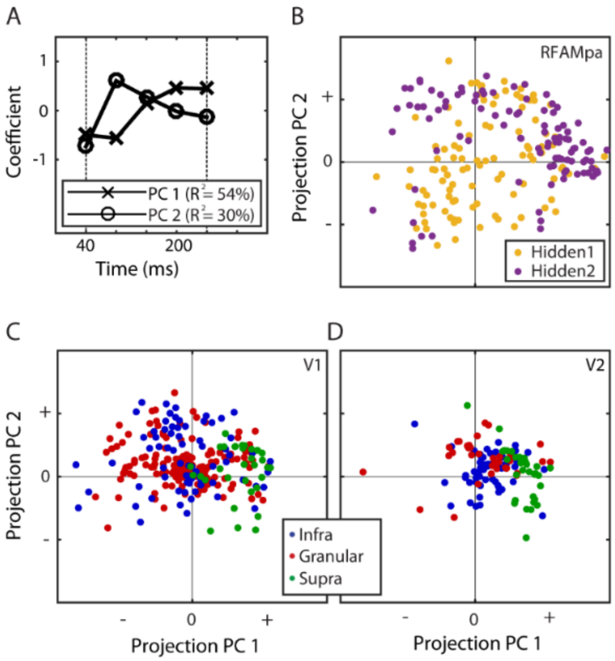

Figure 21. Principal component analysis of the time courses. A) The first two principal components (PC1 and PC2) explained 84% of the temporal response properties for all units and cells. B-D) Projections of the responses of the units of H1 and H2 (B), V1 (C), and V2 (D). The projections revealed no obvious differences between the hidden units and the V1 and V2 cells and the convex hulls of the data points were largely overlapping, showing that the temporal response properties of the hidden units in H1 and H2 were similar to those of the neurons in V1 and V2. Whereas the projections of the V1 and V2 population of granular and infragranular cells onto PC1 and PC2 were positive and negative, the supragranular cells had primarily positive projections on PC1. This feature was also seen in the H2 population, but not the H1 population. This figure shows that the response dynamics as well as the hierarchical organization in V1 and V2 match those found in the recurrent network model.

We reemphasize that the RFAMpa was tasked only with reproducing the average time course of the V2pa (Figure 19A-B) in response to static MSCT patterns. This target response was primarily selective for the *even* class and had a transient-sustained response profile. The RFAMpa solved this task using hidden units with a wide range of MSCT tuning properties and response dynamics that were quite different from the target output response, but qualitatively similar to the tuning and dynamics of V1 and V2 neurons (see Discussion).

3.3.2.4 Hidden units: motion and orientation tuning properties

The rapid and transient dynamics of the hidden units suggest that they could play a role in the detection of moving patterns. We investigated this by simulating translating and rotating random binary noise patterns at various speeds and directions of motion (Materials and Methods). Figure 22 shows tuning curves for four example hidden units based on the response averaged over 10,000 trials and over the first 200 ms after stimulus onset. The different panels show the tuning for MSCT classes (Figure 22A), translational motion (Figure 22B), and rotational motion (Figure 22C). These example units show the range of motion tuning across the two hidden layers, from virtually no effect of motion (example unit in the first row), band-pass speed tuning (second row), low-pass speed tuning (third row), and high-pass speed tuning (fourth row). Note that this tuning emerged even though the network was never exposed to any moving patterns during the training phase.

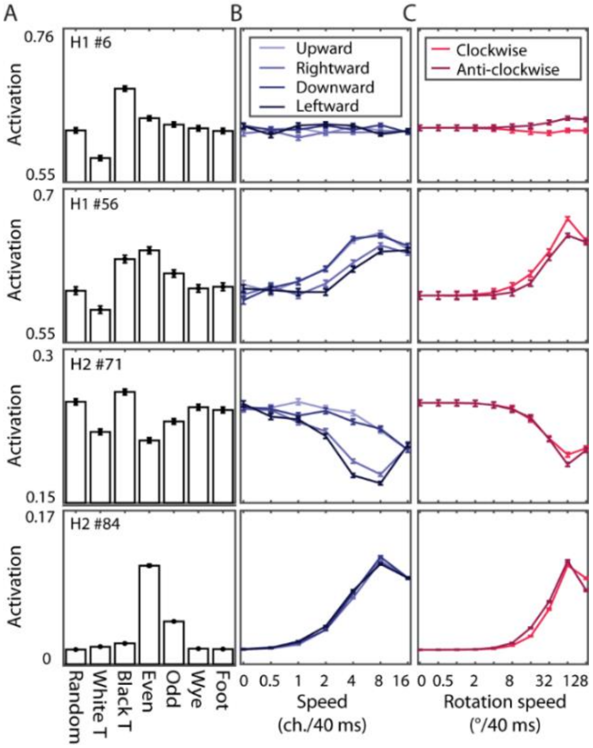

Figure 22. Texture and motion tuning of the RFAMpa hidden units. A) Texture tuning. Mean responses to examples from the six MSCT classes. B) Motion tuning. Response to translating binary random noise patterns at different speeds (x-axis) and directions (legend). C) Same as B, now for rotational motion. Error bars indicate standard error over examples. Note that the static stimuli in panel B and C are identical to the random textures in panel A. This figure shows that, although RFAMpa was trained only on static MSCT and only to reproduce the V2 population average responses in Figure 21A, it self-organized hidden units with diverse selectivity for dynamic stimuli.

An interesting clue about the computations performed by the network comes from comparing the MSCT selectivity (the true goal of the network) to the motion and orientation tuning strength (emergent properties of the network). For each hidden unit we calculated the speed tuning response for translation and rotation (SRt, SRr), direction response (DRt, DRr), and orientation response (OR) by mapping the units' tuning curves in simulated experiments. We then compared the response for any of these stimulus dimensions to the texture response (TRx) (see Materials and Methods).

Figure 23A shows the relationship between SRt and the TRe for all units of H1 (left panel) and H2 (right panel). This shows that the response for the *even* textures was positively correlated with the speed tuned response (H1 r=0.75, H2 r=0.79). Note that most units with an increased/decreased response for the *even* textures (compared to the *random* textures) had an increased/decreased response for the dynamic stimuli (compared to the stationary stimuli), respectively. This is consistent with the examples of Figure 22; hidden unit #6 had a weak speed selective response (small SRt and SRr) and selectivity for *even* textures (near zero TRe) while hidden unit #56 in H1 and #84 in H2 had a strong positive SRt, SRr, and TRe and hidden unit #71 a strong negative SRt, SRr, and TRe selective response (see Discussion).

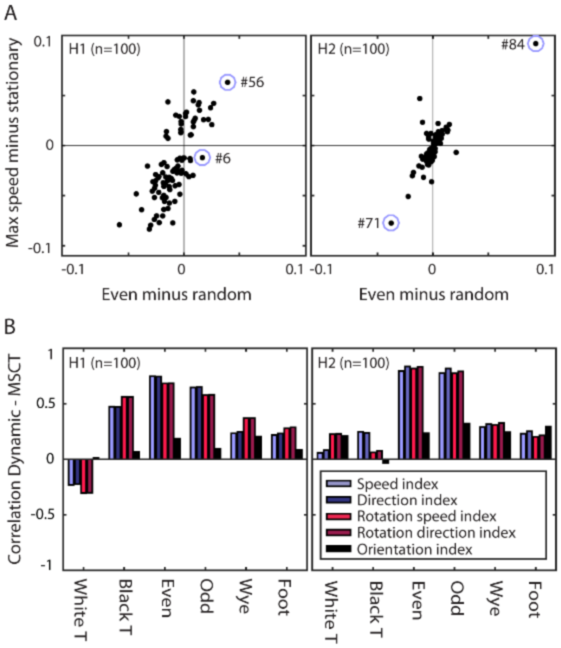

Figure 23. Texture and motion tuning are correlated in RFAMpa hidden units. A) Speed selective responses versus even texture class responses in H1 (left) and H2 (right). The correlation was r=0.75 for H1 and r=0.79 for H2. Open circles indicate the four example hidden units used for Figure 22. B) Correlation between texture responses for each of the six MSCT classes, the motion response, and the orientation response in H1 (left) and H2 (right). This figure shows a strong correlation between motion tuning and texture tuning for the visually salient four-point MSCT classes.

Figure 23B shows the correlation between each of the motion and orientation tuning responses and each of the texture tuning responses, separately for H1 (left panel) and H2 (right panel). This figure shows two important results. First, units with strong orientation tuning did not necessarily have strong four-point MSCT responses (correlations H1 r=0.19, H2 r=0.23). Second, velocity tuning (both speed and direction) was highly predictive of selectivity for the visually salient four-point textures (combined H1 r =0.7, H2 r=0.81) but not the visually salient three-point textures (combined H1 r=0.14, H2 r=0.14) nor the non-salient four-point textures (combined H1 r=0.22, H2 r=0.26). Taken together, these data show that texture tuning, and particularly the selectivity for four-point correlations and motion tuning are closely linked in the RFAMpa (Discussion).

3.3.2.5 Robustness

As the training algorithm of the artificial neural network includes a random initialization of network connectivity, and because the backpropagation algorithm is not guaranteed to find a globally optimal solution, one might be concerned that the variety of response dynamics across the network (Figure 19) and the emergent property of motion tuning (Figure 22) could be an artifact of the training algorithm. To address this, we repeated the full training procedure, with randomly chosen weight initializations 10 times and performed the same analyses as above for each of those networks. In short, this analysis showed that the findings reported above are robust.

Specifically, we performed PCA analysis on the full set of (579x5) time courses. The first two principal components of this data set were very similar to the components shown in Figure 21 (average correlation r=0.84), and the convex hull of each of the 10 trained networks overlapped on average 80% with that of the other networks. The same was true for motion tuning which was correlated with selectivity for the 4th order multipoint correlations in all 10 networks (r between 0.56 and 0.75 for H1 and 0.8 and 0.87 for H2).

Together with the high level of performance on the test set, this shows that the diversity in temporal dynamics, and the presence of significant motion tuning are not artifacts of the random network initialization or suboptimal solutions found by backpropagation, but a robust and salient aspect of how this recurrent network model generates selectivity for multipoint correlations.

Another aspect of robustness is that qualitatively similar network properties were found in the full RFAM network model (trained on all 123 V2 single cell responses). As discussed above, focusing on the population average response in the RFAMpa network had several advantages, but we found analogous properties of the hidden units of the RFAM that reproduced the responses of all 123 recorded V2 cells. For completeness, we list them briefly here. First, both hidden layers had units with diverse MSCT selectivity. Second, the hidden units had complex time courses that could largely be explained by the first two PCs shown in Figure 21 (r=0.87). Third, many hidden units of both hidden layers were tuned to dynamic stimuli and their motion tuning strength was highly correlated with their selectivity strength for the visually salient four-point textures (combined H1 r=0.67, H2 r=0.97) and not the visually salient three-point textures (combined H1 r=0.14, H2 r=0) nor the non-visually salient four-point textures (combined H1 r=0.2, H2 r=0.3). This demonstrates another form of robustness of our results; the self-organized tuning properties of the hidden units occur equally in a network trained to reproduce each of the V2 neurons

(RFAM), or a network trained to reproduce only the average supragranular V2 MSCT response (RFAMpa).

3.4 Discussion

We developed a novel, four-layer recurrent network model to explain the response properties of V2 neurons to local features. This model captured texture tuning, generalized to new stimulus examples from the texture classes, and reproduced not only the mean firing rate, but also the temporal dynamics of the neural responses.

Analyzing the hidden units of the RFAM revealed that texture tuning was more pronounced in the second hidden layer than the first hidden layer, analogous to the difference between V2 and V1, respectively. The dynamic responses of the hidden units were highly diverse but quantitatively similar to those observed in V2 and V1. The complex and diverse dynamics led to the novel insight, and experimental prediction, that the detection of complex form and motion should rely on the same early visual neurons.

3.4.1 Comparing modeling approaches

Our first attempt to model the neural responses made use of a standard and rather general approach that seeks to capture the responses as a sum of linear-nonlinear filters (Pillow & Simoncelli, 2006; Rust et al., 2005). While an acceptable fit to training sets could be obtained, predictions of the fitted models failed to generalize – i.e., they did not properly predict responses to stimulus examples outside of the training set. Thus this modeling procedure did not capture the essence of the computations, or insights into a possible mechanism.

There are a number of incremental changes one could make to the LN approach to improve its generalization performance such as adding higher-order filters, estimating non-separable high-dimensional nonlinearities, estimating space-time instead of space-only filters, or adding a second stage in which filters are combined nonlinearly (Rust et al., 2005). However, given that the simple LN model already captured the training data well, it seems likely that these additions – which add significant complexity but retain the core structure and parameters of the LN model – would merely increase overfitting.

We believe this to be an important general point. A reverse-correlation analysis can always be expressed in terms of a stack of linear-nonlinear channels that capture some fraction of the variance in an experimental data set, and - given a sufficient number of channels - one can approximate any transformation. Often the filters provide an intuitive way to understand the input-output mapping (e.g. oriented filters for neurons with orientation selectivity (Ts'o, Gilbert, & Wiesel, 1986) or space-time oriented filters for neurons with motion tuning (Rust et al., 2005)). However, there is no guarantee that this is the case and our analysis warns against a mechanistic interpretation of such filters. Filters are informative only if they generalize to new examples from the same class (e.g. other oriented patterns, other moving patterns) or if they generate novel predictions that can be confirmed experimentally (Rust et al., 2005). Without such confirmation of generalization, the model cannot suggest insight into the underlying computational mechanisms.

The RFAM, on the other hand, generalized well out-of-sample. The RFAM approach differs from the LN approach in many ways, making it difficult to isolate the reason for their contrasting performances. Nevertheless, it is instructive to consider which factors contributed to the better generalization in the RFAM approach.

First, we trained the RFAM network on the full time course while we used only the mean firing rate to determine the LN model parameters. While one could extend the LN model with spatiotemporal filters (as has been done in previous work), estimating space-time filters would lead to even worse overfitting – as it would add free parameters more rapidly than it would add constraints. Here, because the stimuli were all unmodulated in time, and the responses across stimulus categories are dynamically similar within each neuron (Yu et al., 2015), even the restriction to space-time separable filters would suffer from this problem. In a recurrent network, however, the intrinsic dynamics predict a time course and adding time points to the to-be-explained data set increases the constraints on the model without increasing the number of free parameters. These additional constraints reduce the tendency to overfit the data.

Second, we trained a single RFAM network to generate the output of all V2 neurons simultaneously, whereas the LN approach determines a separate, independent filter for each V2 neuron. Forcing a set of hidden units to generate a representation that results in well-matched output of all V2 neurons likely reduces overfitting the noise in the response of any single V2 neuron. Incorporating this approach in the LN model would lead to a feedforward network with a single hidden layer and an output layer representing, for instance, all V2 neurons.

This brings us to a final important point about comparing models. The universal approximation theorem (UAT) for feedforward networks with a single hidden layer (Hornik et al., 1988) proves that a feedforward network exists that can perform just as well as the RFAM (or a single layer recurrent network, which is also a universal approximator (Funahashi & Nakamura, 1993). In other words, goodness of fit, or lack of such fit, cannot be taken as evidence to support the need for recurrent connections, nor the need for two layers. In fact, the choice among these network architectures can never be based on the performance of the network alone. Instead, such choices must be based on other aspects of the modeling approach.

First, there are practical matters such as the ease with which a solution can be found in a specific architecture (the UAT guarantees that a solution exists, but there are no algorithms that are guaranteed to find this solution). Second, a feed-forward network maps input sequences to neural responses by using spatiotemporal weights that allow each neuron to look back in time to previous inputs. This can be a convenient short-cut to capture neural responses but it punts on the mechanistic question how a network integrates information over time. If this question is of interest, one has to look beyond feedforward networks. Third, a-priori knowledge such as the ubiquity of recurrent connections in the brain can motivate one model over others.

These considerations motivated us to develop the novel RFAM approach, instead of incrementally adding to the LN approach. The true value of the RFAM approach, however, is not that it captures the data better (many models could do that), but that it leads to a novel mechanistic hypothesis of the computations underlying higher-order form processing (Figure 21, Figure 22), and testable predictions about the relationship between form and motion processing in early visual cortex (Figure 23).

3.4.2 Form and Motion

Why would motion and form analysis go hand-in-hand? Motion detectors can be characterized as logical-and operations: a moving object was here at this time *and* there some time later. Four-point correlations can similarly be detected as the logical-and of two orthogonal orientations. Consistent with this, many V2 neurons appear to have sensitivity to orthogonal orientations (Anzai et al., 2007). As our analysis of the LN model shows, however, feedforward solutions in which the logical-and is computed using high thresholds do not generalize well across the textures in a class.

We, therefore, propose that recurrent connections provide a robust way to compute a logical-and (Salinas & Abbott, 1996) while also providing a rudimentary memory that allows the comparison of neural output at different times (Joukes et al., 2014). The duration of this memory, or the effective integration time of (parts of) the network, can be adjusted by the strength of the recurrent connections. This flexibility allows the network to detect first- and second-order statistics in one part of the texture and compare this with first- and second-order statistics in one or more other parts of the image after a short delay. For images that are presented abruptly and then remain static during the delay, this comparison will yield selective responses to specific third- and fourth-order spatial statistics. For images that translate in time, this will yield sensitivity to motion patterns, including those driven by high-order statistics (Chubb & Sperling, 1988; Clark, Bursztyn, Horowitz, Schnitzer, & Clandinin, 2011). This leads to our prediction that motion and texture tuning are intricately entwined.

At face value this claim appears to be at odds with the view that form and motion processing proceed along largely independent pathways in the brain (Hubel & Livingstone, 1987; Livingstone & Hubel, 1984). However, such claims are typically based on the lack of correlation between tuning for orientation and tuning for motion. This correlation is also low in the hidden units of the RFAM network (H1 r=0.31, H2 r=0.37), but orientation tuning is only one aspect of form selectivity: our analysis predicts specifically that selectivity for four-point correlations should be most strongly correlated with motion tuning (Figure 22 and Figure 23).

Beyond the level of tuning for oriented bars, a link between motion and form tuning does find substantial experimental support in the interactions between complex shapes and motion in early and mid-level visual areas (Kourtzi & Kanwisher, 2000; Kourtzi, Krekelberg, & van Wezel, 2008; Krekelberg, Dannenberg, Hoffmann, Bremmer, & Ross, 2003; Krekelberg et al., 2005). In addition, anatomical evidence shows a significant degree of convergence of form and motion processing in V1 (Callaway & Wiser, 1996; Fitzpatrick, Usrey, Schofield, & Einstein, 1994; Sawatari & Callaway, 2000) as well as V2 (Sincich & Horton, 2002). Nevertheless, the model's specific prediction of a link between motion sensitivity and form analysis await a direct experimental test.

3.5 CONCLUSION

A network with two recurrently connected hidden layers captured the selectivity of V1 and V2 neurons for multipoint correlations and generalized to new examples from the texture classes. Analysis of this network strongly suggests that visually salient four-point correlations can be detected by a network with diverse selectivity for all visually salient MSCT texture classes and with complex time courses that closely match the properties of V1 and V2 neurons. In this network, many units were motion tuned and the extent of motion tuning was correlated with tuning for the visually salient spatial multipoint correlations. This leads to the novel prediction of a specific overlap between tuning for complex form and motion in early visual processing.

More broadly, our work shows that recurrent connectivity – a defining characteristic of all cortical networks – can solve computational problems in unexpected ways. We trained an artificial recurrent neural network to capture the full time course of the neural response to a sensory input and, in doing so, uncovered a new neural solution to a complex computational problem. Because the artificial network can be probed in depth and at length, it lends itself well to generate novel and experimentally testable predictions. We believe that this approach is a useful method to uncover novel computational principles well beyond early visual cortex.

4 SPACE-TIME INTEGRATION BASED ON RECURRENT NETWORK DYNAMICS IN PRIMARY VISUAL CORTEX

4.1 INTRODUCTION

In chapter 2 we discussed the recurrent motion model (RMM), an artificial recurrent neural network that we fitted to the motion tuned responses of MT cells (Joukes et al., 2014). The RMM provided the novel insight that the MT motion tuning properties cannot be fully captured by 1st and 2nd order space-time correlations. Instead, the recurrent connections of the RMM integrated the motion input stimuli over more than two time steps (monitor frames) in a nonlinear way. This is in line with previous findings that MT cells typically become more direction tuned over more than two successive motion steps (Mikami, 1992) and the behavioral phenomenon of sequential recruitment in a human direction discrimination performance task (McKee & Welch, 1985).

In chapter 3 we discussed the recurrent form analysis model (RFAM), an artificial recurrent neural network that we fitted to a set of recorded V2 neurons that were selective for multi-point spatial correlation textures (MSCT). The RFAM captured the V2 MSCT selectivity by integrating the responses of multiple (hidden) units, each with a selectivity for a small subset of patterns per MSCT class. With differing spatial RFs, the integration over multiple hidden units and time steps resulted in sensitivities for moving patterns. This led to the surprising RFAM prediction that MSCT selectivity (static higher order spatial correlations) overlaps with motion tuned (space-time correlations) responses (Joukes et al., 2014).

The emerging properties of the RMM and the RFAM showed that our ANN approach provides a powerful toolset for a detailed investigation of how a neural system computes. However, the MT and V2 data sets had several limitations that prevented us to test the model predictions with the experimental data set that we used to fit the models. First, both data sets were collected with cell recordings in which each neuron was lost after the recording session and could, therefore, not be used to directly test the model predictions. The second limitation of the data sets was that most neurons were recorded independently. A fit of a single RMM or RFAM with multiple output units corresponding to our selection of MT or V2 cells, therefore, wrongfully assumed that all responses were recorded at the same time. For instance, the MT and V2 neurons had variable RF sizes and we modeled this with a single network by taking the average RF size over all cells. This approach, of course, cannot capture potential RF size dependent tuning properties. Third, the large (average) RF size of MT neurons forced us to collapse the 2-dimensional experimental motion stimuli into 1-dimensional motion simulations. Therefore, potential MT tuning properties specific to 2-dimensional motion could not be captured by the RMM. Finally, for the MT data set it was practically impossible to collect enough experimental data for a successful fit of the RMM to the experimental motion stimuli and MT responses, even for 1-dimensional motion stimuli. Instead, we assumed pattern invariance for all our MT neurons (Albright, 1984). We were able to overcome this limitation for the RFAM and this resulted in a much closer link between model and experimental data and, therefore, resulted in more credible model insights and predictions.

In this chapter we present a model - the recurrent space-time model (RSTM) - where we improved upon all previously discussed limitations by taking full advantage of chronically implanted multi-unit floating microelectrode arrays (FMA) in primary visual cortex (V1) of the awake behaving macaque. First, the same neurons can be recorded by the FMA over several recording sessions, sometimes spanning weeks or even months. This provides the opportunity to fit an artificial neural network to the experimental data, investigate the properties of the fitted network, and test the model predictions a few days later on the same neurons. Second, the FMA consists of 32 electrodes that simultaneously record the activity of a population of V1 neurons. Therefore, a successful fit of multiple ANN output units to multiple FMA electrodes allows for an investigation of network effects that rely on dependencies between neuronal responses to the same experimental stimulus. Third, V1 has reduced RF sizes compared to MT or V2. In contrast to the RMM and MT, we can, therefore, explore V1 tuning properties in high space-time resolution. Finally, the reduced

RF size of V1 neurons also allows for the collection of a large enough experimental data set to fit an ANN directly to experimental stimuli and recorded neuronal responses.

We investigated V1 space-time tuning properties following a four step approach that spanned four consecutive days. First, on day one we stimulated the combined RF of the neurons recorded by the V1 FMA with 2-dimensional dynamic white noise stimuli. This allowed us to extract the V1 tuning properties in a data-driven way. Second, we fitted the RSTM to the noise stimuli and the recorded V1 responses. Third, a detailed investigation of the RSTM with simulations of white noise, orientation, texture, and motion stimuli provided us insights into the tuning properties of the model output units. Fourth, we presented a selection of the most salient stimuli to the monkeys with a V1 validation experiment to test the RSTM predictions three days after the white noise experiment.

4.2 MATERIALS AND METHODS

4.2.1 Electrophysiology

Neuronal activity was recorded with 32-channel Floating Microelectrode Arrays (FMA) that were implanted in parafoveal primary visual cortex (V1) of two adult male macaques (Macaca mulatta). Monkey M had one array in the right hemisphere with receptive fields centered at x: 2.8°, y: -5.4 ° from the fovea; Monkey Y had two arrays in the left-hemisphere – FMA1 with receptive fields centered at x: -3.2°, y: -3.2°, and FMA2 with receptive fields centered at x: -1.45°, y: -1.5° from the fovea (for the current experiments only FMA2 was used). The experimental and surgical protocols were approved by the Rutgers University Animal Care and Use Committee, and were in agreement with the National Institute of Health guidelines for the humane care and use of laboratory animals. The neural signals were amplified, filtered and sampled at 30 kHz using a Grapevine neural interface system and Trellis software (both produced by Ripple). Wavelet transforms were used for spike detection (Nenadic & Burdick, 2005) and Klustakwik cluster analysis for spike sorting (http://klusta-team.github.io/klustakwik/).

4.2.2 Visual stimulation

The visual stimuli were computed in Neurostim (http://neurostim.sourceforge.net), and displayed on a 20-inch CRT monitor (Sony GDM-520). The display was 40° x 30°, with a resolution of 1024 x 768 pixels, had a refresh rate of 150 Hz (set to 75 Hz for the current experiments, ~13 ms per frame), and a mean luminance of 30 cd/m2. The monkeys viewed the stimuli from a distance of 57 cm in a dark room (<0.5 cd/m2) while seated in a standard primate chair (Crist Instruments, Germantown, MD). Eye position was recorded with an infrared tracker at a sampling frequency of 250 Hz (EyeLink2000; SR Research). Trials in which eye position deviated from a 2° wide square window centered on the fixation spot were excluded from analysis.

4.2.3 Stimuli and experimental paradigm

We investigated the V1 space-time tuning properties following a four step approach that spanned four consecutive days. On day one we stimulated the combined RF of the neurons recorded by the V1 FMA with 2-dimensional dynamic white noise stimuli. This allowed us to extract the V1 tuning properties in a data-driven way. After the conclusion of the white noise experiment we fitted an artificial recurrent neural network to the noise stimuli and the recorded V1 responses. Next, a detailed investigation of the fitted neural network with simulations of several stimulus classes provided us the tuning properties of the model output units. We then tested the predictions of the model on day four with a validation experiment where we presented the most salient simulated stimuli to V1.

4.2.3.1 White noise experiment

The stimulus was centered on the combined RF of the cells recorded by the FMA. It appeared 250 ms after the monkey fixated a central red dot at the center of the screen. The monkeys maintained fixation for 3.6 seconds and received a reward after each successful trial.

Trials consisted of nine 240 ms (75Hz, 18 monitor frames) pattern streams with a new 2-dimensional (22x22 values) noise pattern per monitor frame. Each pattern stream was followed by a 160 ms blank period during which a grey screen (luminance 30 cd/m2) was shown. The noise patterns were low pass filtered by convolving with a Gaussian ($\sigma = 1.5$ noise value). We maximized contrast per monitor frame by scaling the noise pattern to the maximum range of the monitor (see Figure 25 for an example pattern stream). A total of 100 unique noise pattern streams were presented in randomly interleaved trials that were repeated 7 (monkey M) and 10 (monkey Y) times.

4.2.4 Model fit of the experimental white noise data

4.2.4.1 Time-delay artificial recurrent neural network

We fitted an artificial recurrent neural network – the recurrent space-time model (RSTM) - to the experimental white noise stimuli (input) and the recorded V1 response dynamics (output). The RSTM was implemented in the Matlab Neural Network Toolbox (version 8.4) and was based on (an improved implementation of) the Elman network (Elman, 1990). Units in such an artificial network are considered a crude approximation of a neuron or a group of neurons (Figure 25B). The units were interconnected with adjustable weights simulating synaptic connections with variable strength. Each unit also had an adjustable bias value. The network had one input, one hidden, and one output layer. The input layer consisted of 484 units that each simulated one of the unique 22x22 low pass filtered white noise values of the experimental noise stimuli. The input layer was fully connected to the first hidden layer in a feedforward manner. The hidden layer had 100 units that were fully connected to the output layer of the RSTM in a strictly feedforward manner. In addition, the hidden units were laterally/recurrently connected to all hidden units within their layer. The recurrent input of the hidden layer was time-delayed by one simulation time step compared to the feedforward input originating from the input units.

The goal of the RSTM was to capture the response of 32 V1 FMA electrodes in an output layer with 32 units. We fitted separate RSTMs to the MUAe and the isolated single cells. The output for each unit (i) was calculated by first determining the weighted sum of its inputs plus the bias value: $\mathrm{x_i} = \sum_{k}^{i} \mathrm{w_{ik} y_k} + \mathrm{b_i}$, where the index k runs over all units that are connected to unit i, and then passing this through a transfer function: $y_i = \ 2/(1 + e^{-2*\mathrm{yi}} - 1)$. for the hidden units and a sigmoid transfer function for the output units: $y_i = \ 1/(1 + e^{-\mathrm{x_i}})$.

4.2.4.2 Output pattern streams

The RSTM was trained to reproduce the V1 FMA responses. In the language of artificial neural networks these are called the target patterns, or because they are the responses of the output units, output patterns (Figure 24). We applied the following preprocessing steps to the V1 FMA responses in preparation for the RSTM output patterns. First, per ms, we corrected for outliers in the MUAe by removing the highest two and lowest two values over repeats. Second, we binned the MUAe and firing rate responses in 13 ms time bins (the monitor frame rate). Third, we removed the stimulus independent time course of the response by subtracting out the mean over all trials and repeats per 13 ms time bin (Figure 24C). Fourth, to account for the difference between instantaneous processing of the artificial network and delayed response dynamics of the V1 neurons, we shifted the response forward by 53 ms (the average onset latency of our recorded V1 cells). Next, we separated the response to the nine noise pattern streams per trial, creating a total of 900 output pattern streams of length 18 (=240 ms/13 ms). Finally, we scaled the output pattern streams to a suitable range for the artificial neural network by dividing the responses by three standard deviations over all responses of all electrodes. Note that the network activation, therefore, is defined in terms of standard deviation over all responses of all electrodes. Moreover, due to the subtraction of the stimulus independent time course, the remaining signal deviation from zero reflects primarily stimulus dependent responses.

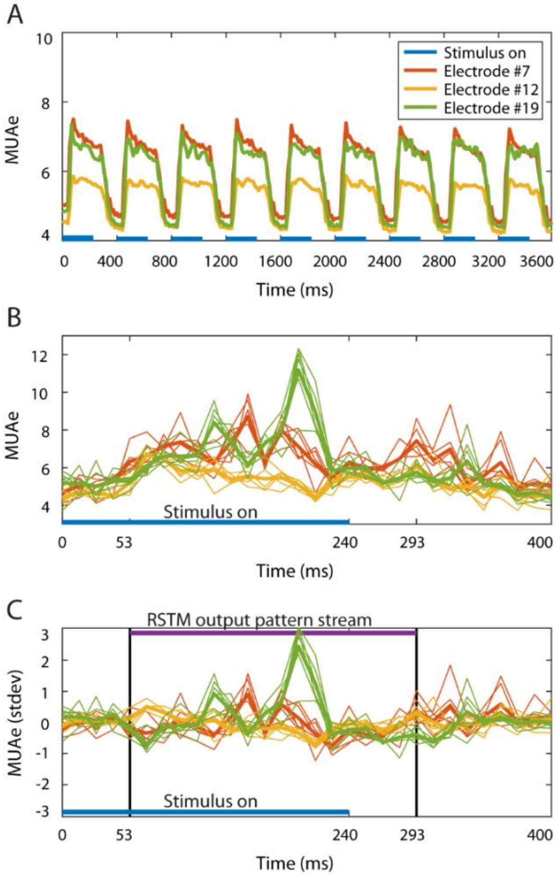

Figure 24. MUAe of three example V1 electrodes in response to white noise patterns. A) Response of three electrodes (see legend) to 3.6 second trials (mean over 100 trials and 7 repeats). Each trial consisted of nine 240 ms (18 monitor frames) white noise pattern streams (blue lines represent stimulus-on phases), each followed by a 160 ms blank period. The figure shows a strong stimulus driven MUAe of the three example electrodes that moved back to baseline during the blank period in between the stimulus pattern streams.

B) Response to the presentation of one white noise pattern stream for the three example electrodes shown in (A). The figure shows a highly dynamic, variable response that was consistent over the seven repeated presentations of the same white noise pattern stream. C) Target pattern streams for the RSTM output units. The stimulus independent time course of each V1 FMA electrode response was removed by subtracting the mean response over all trials and repeats per 13 ms time bin. The V1 neurons had an average onset latency of 53 ms; the target output patterns (purple box) were corrected for this.

4.2.4.3 Input pattern streams

The RSTM input pattern streams matched the experimental white noise stimuli one-to-one with the only adjustment that the scaling was changed from the maximum luminance range of the monitor to a more suitable range for the artificial network (-1 and 1). Similar to the output pattern streams, the trials were separated into nine independent pattern streams, resulting in a total of 900 input pattern streams of length 18 (240 ms/13 ms).

4.2.4.4 Training phase

Before training the networks, we initialized the weights and bias values of all layers with the Nguyen-Widrow algorithm. We trained the recurrent neural network on the input and output pattern streams in the following way. First, we randomly chose one of the input pattern streams and presented this to the network. Second, we calculated the response of all units in the network for 18 time steps. Third, we randomly selected one of the repeats as the network target for this training cycle (epoch). Fourth, we calculated the error as the mismatch between the response of the RSTM output units and the V1 targets. Finally, this error was then used to modify the connection weights in the network using error back-propagation-through-time.

To prevent overfitting, we used (all repeats of) 90% of the input and output pattern streams to fit the RSTM (train set) and 10% to monitor out-of-sample generalization performance (test set). The error over time showed that a training period of four million epochs was sufficient to reach convergence for the train set (i.e. further training contributed little to a reduction in error) without a worsening of the error for the test set (Figure 26A-B). We fitted separate RSTMs to the MUAe (RSTM MUAe) and firing rate (RSTM Spike) data sets for Monkey M and Monkey Y. Network parameters were then frozen and we investigated the trained network.

To estimate the performance of the RSTM output units as a measure of the level of signal contained in the responses of each electrode for the train and test set and for the V1 validation experiments we will discuss in detail below, we define the consistency of a response as the mean square of the correlation between that response and the response observed in identical repeats of the same stimulus sequence (see similarly colored lines in Figure 25A). We used this measure to define the performance of each RSTM output unit for a single sequence as the ratio of the R^2 between the model response and the neural response for a single repeat divided by the consistency of the response over repeats (Figure 26, Figure 32, Figure 33, Figure 34, Figure 35). In other words, the consistency measure shows how much variance in the V1 response can be explained or predicted by the model units as a fraction of how much of the variance a V1 neuron can explain for its own response to a repeated presentation of the exact same stimulus sequence. To quantify overall performance, these normalized correlation values were averaged over stimulus sequences.

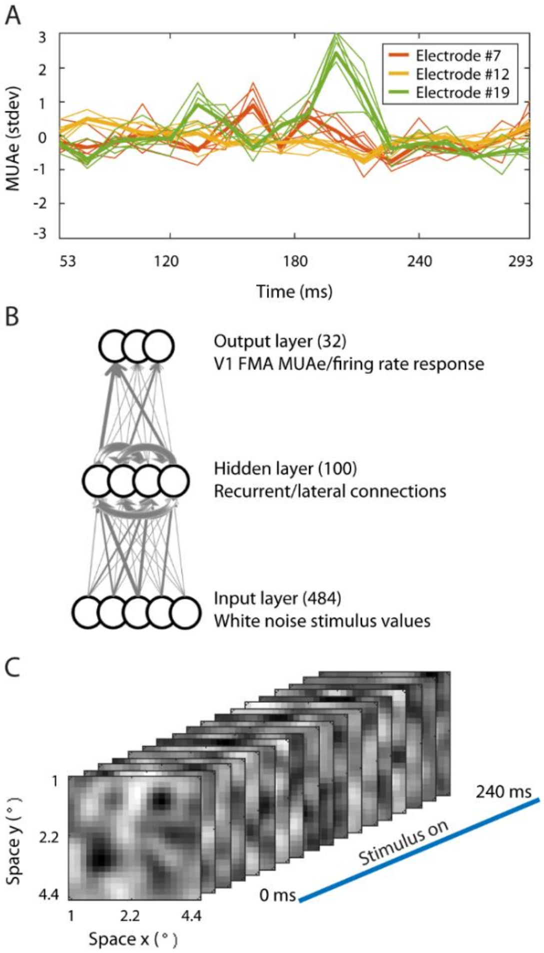

Figure 25. Training the recurrent space-time model (RSTM). A) Example output pattern streams for three RSTM output units (based on the MUAe of the three example V1 FMA

electrodes shown in Figure 24. Note that the response difference to the same repeated input pattern sequences (similarly colored lines) indicates the consistency of the response that we use throughout the results section as a baseline for network performance. B) The RSTM had 484 input units, one for each experimental white noise stimulus input value (see panel C). The input units were all-to-all connected to the units of the first hidden layer. The hidden layer had 100 units which were recurrently connected to all units within the same layer (thick gray lines). They, in turn, were all-to-all connected to the 32 output units. C) Example RSTM input pattern stream from the white noise experiment. A stream of 18 low pass filtered white noise values (22 by 22 values), one for each monitor frame at 75Hz, was presented to the network. The goal of the network was to adjust the weights of the connections between the units in an iterative procedure (backpropagation through time) to simultaneously reproduce the output of each of 32 V1 FMA electrodes (MUAe or firing rate) in the 32 output units.

4.2.5 Simulations of the fitted artificial recurrent neural network

4.2.5.1 Linear-nonlinear model

A linear-nonlinear (LN) model (Chichilnisky, 2001; Rust et al., 2004; Simoncelli et al., 2004) describes neural responses as a set of linear space-time filters and their corresponding nonlinearities. We followed the same steps as previously discussed for the RMM and RFAM. We used the spike triggered average (STA) and the spike triggered covariance (STC) methods to estimate these filters (Chichilnisky, 2001; Rust et al., 2004; Simoncelli et al., 2004) for the RSTM output units using ten million newly generated white noise stimuli with identical parameters as the experimental white noise pattern streams described previously. Based on the STA and STC we estimated the information captured by the maximally informative filters using the iSTAC method (Pillow & Simoncelli, 2006). We also determined the nonlinearity associated with each filter by dividing the histogram of the projected spike triggered ensemble by the histogram of the projected raw stimulus ensemble, over four standard deviations away from the mean. For an example RSTM output unit the 16 most informative filters and their nonlinearities are shown in Figure 27A-C.

4.2.5.2 Interactions over hotspots

To highlight sensitivity of the RSTM output units to multiple spatial locations within the full RSTM input space, we define the space-filter (Figure 27B) as the square root of the squared and summed filter over time and we define the sensitivity filter as the square root of the 16 most informative squared and summed space-filters (Figure 27E and Figure 28). The sensitivity filters of the RSTM output units revealed that individual units were sensitive to multiple distinct locations in space (‘hotspots’, see Figure 28). We determined tuning and interaction effects within and between hotspots with orientation, texture, and motion stimuli in the following way. First, we estimated the location and size of hotspots (RF1 and RF2) for Monkey M and Monkey Y (11x11 and 9x9 input units, respectively) that were shared by most of the RSTM output units. We then presented the stimuli (described in detail below) to RF1 (with all other inputs set to zero), RF2 (with all other inputs set to zero), or the full RSTM input space (RFFull). Finally, we presented all stimulus condition combinations to RF1 and RF2 at the same time (RF12). We defined nonlinear interactions over RF1 and RF2 of the RSTM output units as the difference between the response to RF1 + RF2 and the response to RF12.

4.2.5.3 Orientation tuning over hotspots

We created 1000 oriented stimuli using 1-dimensional white noise values in the x-direction replicated in the y-direction with a number of values matching the size of the RF condition (22, 11, and 9 for RFFull, RF1, and RF2 for Monkey M and Monkey Y, respectively). We rotated these 2-dimensional patterns with one of 19 angels between 0° and 180°. Finally, we low-pass filtered the oriented patterns with a Gaussian filter (σ = 1.5 input value). The same pattern was presented to each output unit of the model for five time steps (67 ms) and for all RF location conditions described previously. The mean response over 1000 patterns and five time bins gave the orientation tuning curves (mean orientation tuning) per RF location condition (see example RSTM output unit Figure 29B-C).

4.2.5.4 Texture tuning over hotspots

The texture stimuli were identical to the experimental and model stimuli we previously discussed in Chapter 3. The stimuli were checkerboard arrays that consisted of black and white checks with the number of values matching the size of the RF conditions. Checkerboards were either random (check colors assigned independently and with equal probability to black or white), or constructed to contain only spatial correlations of a specific spatial configuration and order. We refer to the latter as multipoint spatial correlation textures (MSCT). We simulated four MSCT classes: two classes contained visually salient three-point correlations (*white triangle* and *black triangle*), two classes contained visually salient four-point correlations (*even* and *odd*). Stimuli from these four classes were generated via a Markov recurrence rule (Victor & Conte, 1991, 2012). It is important to note that for each MSCT, the specific multipoint correlations are fixed, and there are (on average) no correlations of lower orders (e.g. the *even* stimulus class has a specific fourth-order correlation, but does not have first- (mean luminance), second- (power spectra/spatial frequency content) or third-order correlations). Put differently, these classes form a basis to study the influence of each kind of multipoint correlation.

We simulated 5000 low-pass filtered (with a Gaussian filter with σ = 1.5 input value) patterns per texture class and RF condition, static over five time bins (67 ms). The average response over patterns and time to the four MSCTs with the response to the *random* textures subtracted (mean MSCT selectivity), gave us the selectivity of the RSTM output units to each MSCT class (see Figure 30A-B for MSCT selectivity of an example RSTM output unit and Figure 30C for example texture stimuli).

4.2.5.5 Motion tuning over hotspots

The motion tuning properties of the RSTM output units were calculated with a simulation of 1000 low pass filtered (by convolving with a Gaussian with σ = 1.5 noise value) 2-dimensional white noise patterns (identical to single frames of the white noise stimuli) that were shifted with one of five speeds (3.75, 7.5, 15, 30, or 60 °/s) in one of eight equally spaced directions (*rightward*, *rightward-downward*, *downward*, *downward-leftward*,

leftward, *leftward-upward*, *upward*, or *upward-rightward*) over five time steps (67 ms). Tuning curves based on the mean response over patterns and time (mean motion tuning) were generated for each RSTM output unit and RF location condition (see Figure 31 for an example RSTM output unit).

4.2.6 V1 validation experiment

We tested the predictions of the RSTM output units with a V1 validation experiment three days after the white noise experiment. Per RSTM output unit and stimulus condition (orientation, texture, and motion) we first determined the preferred and non-preferred tuning for each stimulus condition. Next, we selected the patterns at the preferred and non-preferred tuning that induced the highest and lowest response, respectively. We selected eleven patterns per stimulus condition. Eight patterns based on the highest at the preferred and lowest at the non-preferred tuning for RF1 (1-2), RF2 (3-4), RFFull (5-6) and RF12 (7-8). We selected three patterns to test nonlinear interactions over RF1 and RF2; we presented the stimulus combinations that resulted in the maximum response difference between the sum of the response to RF1 (9) plus RF2 (10) compared to RF12 (11).

These validation stimuli were selected on the basis of 15 RSTM MUAe output units and 15 RSTM Spike output units that had the maximum nonlinear interaction summed over orientation, texture, and motion conditions. Figure 29 and Figure 30 show the selected stimuli for orientation and texture stimuli for two example RSTM output units. Note that, even though we selected 11 stimuli per stimulus conditions for these 30 selected output units, all FMA electrodes recorded the response to all selected stimuli. In other words, a total of 30 (15 RSTM MUAe and 15 RSTM Spike output units) times 11 (RF location conditions) times 3 (stimulus conditions) times 64 (32 MUAe and 32 single cells) data points (=990) were available for analysis.

For the V1 validation experiment, the stimuli were scaled back to the maximum luminance range of the monitor and centered on the combined RF of the V1 FMAs. The subjects fixated a red central fixation point for 3.6 seconds while a stream of validation stimuli was presented. The duration of each pattern stream was five monitor frames (67 ms) following a blank period of 13 monitor frames (173 ms) to allow the V1 response to return to baseline before a new pattern stream of five frames was presented. The stimuli were randomly interleaved in 66 trials, each with 15 pattern streams, that collectively contained all 990 selected stimuli (30 selected output units times 11 RF conditions times 3 stimulus conditions) and were repeated 11 and 10 times for Monkey M and Monkey Y, respectively. Finally, the FMA MUAe and spike sorted responses were first preprocessed following the same steps as previously described for the white noise experiment. We scaled the response of each electrode (MUAe or firing rate) to the response range of each RSTM output unit with a division of the maximum response of each electrode followed by a multiplication of the maximum response of each RSTM output unit.

4.3 RESULTS

With a four step approach spanning four consecutive days we investigated the space-time response properties of neurons recorded with a chronically implanted floating microelectrode array (FMA) in primary visual cortex (V1) of the awake behaving macaque. On day one we stimulated V1 with dynamic white noise that allowed us to extract response dynamics in high resolution and in a data-driven way. Next, we fitted an artificial recurrent neural network on the experimental data and ran network simulations with various classes of visual stimuli. Three days after the white noise experiment we tested the model predictions with a V1 validation experiment where we presented the most salient predicted stimuli to the V1 neurons (see Materials and Methods). In the next sections we show results for our multi-step approach, starting with the fitting process of an artificial recurrent neural network to the V1 white noise experimental data.

4.3.1 Training the recurrent space-time model

Previously, we showed that artificial recurrent neural networks can capture complex nonlinear space-time interactions of motion tuned neurons in the middle temporal area (Joukes et al., 2014) and selectivity for (static) higher-order spatial correlations of early visual form tuned neurons in secondary visual cortex. This motivated us to fit our experimental data with a recurrent neural network – the recurrent space-time model (RSTM). We fitted the RSTM separately to multi-unit data (MUAe output units) and single unit data (spike output units).

Figure 26 shows the performance of three example RSTM MUAe output units (electrode #12, #11, and #8, Monkey M) per repeat of identical pattern sequences after four million training cycles (panel A) and over training cycles (panel B). Our measure of performance for stimulus sequences was the explained variance of the full time course normalized to the variance that was shared between the response to repeated, identical sequences (consistency V1, yellow bars panel A). A performance of one corresponds to a model unit that explains the variance of the neural response equal to how well an individual response (single repeat) can explain all other responses (other repeats) to the same stimulus sequences (see Materials and Methods). Overall, unit performance was defined as the average over all patterns in either the train set (red) or test set (blue). The RSTM output unit that was fitted to electrode #12 (top panels) captured almost all of the variance of the V1 response while generalizing almost perfectly to out-of-sample noise patterns in the test set. Second, for unit #11 (middle panels), the performance on both train and test set outperformed the V1 consistency for that electrode. This shows that the RSTM output unit captured the noise pattern inputs with a more consistent fit than individual repeats to identical noise input pattern streams. In contrast, for unit #8 (bottom panels), the RSTM failed to capture all of the V1 stimulus dependent response variation.

For the population of 32 RSTM MUAe output units, the R^2 values for the train set closely matched the generalization performance for the test. This shows that, even when an RSTM output unit failed to capture all V1 response variation, the part that was captured generalized well to out-of-sample noise patterns. For the remainder of the results section, however, we will primarily focus on the RSTM output units with high performance on the test set (11 electrodes with cutoff R^2 = 40%).

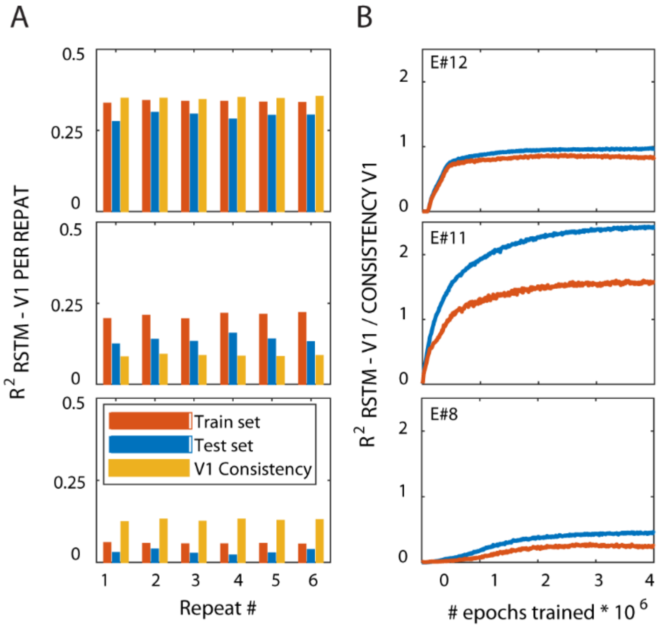

Figure 26. Performance of the RSTM. The network parameters were adjusted iteratively to approximate the MUAe of 32 electrodes or 32 isolated single cells by the 32 output units of the RSTM MUAe or RSTM Spike, respectively. Performance measures the explained variance between the activation of the RSTM output units and the MUAe or firing rates of the target V1 response as a ratio of the consistency of the V1 response (see Materials and Methods). Performance is shown for the train set (red) and the test set (blue). A) Performance per repeat (mean over all pattern sequences) for three example units on the train set (red) and the test set (blue) without adjusting for the V1 consistency (yellow, mean explained variance between the V1 response to a single repeat with all other single repeats). B) Performance over training cycles of the three example units for the train set

and the test set (mean over repeats). This figure shows that some RSTM output units captured a large fraction of the variance in the V1 response that matched the consistency for that V1 neuron (unit #12) and sometimes surpassing the consistency (unit #11). Other units failed to capture the full consistent response dynamics of the V1 responses (unit #8).

4.3.1.1 Linear-nonlinear model of the RSTM output units

The experimental white noise data sets are ideally suited for reverse correlation, a technique that allows for a lower-dimensional description of the complex dynamic V1 responses (Chichilnisky, 2001; Rust et al., 2004; Simoncelli et al., 2004). This technique describes a neuron in terms of an equivalent feedforward linear-nonlinear (LN) model by estimating a set of linear filters and their static nonlinearities (see Materials and Methods). However, in practice, there is not enough time to reliably estimate these filters at a high spatiotemporal resolution.

The close match between the performance of the RSTM output units on the test set of the experimental data (Figure 26), however, suggests that the high performing output units generalize to all white noise patterns. Without the practical limitations of real experiments, we took full advantage of the artificial neural network and estimated the 1st and 2nd order LN models for the RSTM output units with a simulation of millions of noise patterns. Note that the reliability of the estimated LN models is limited by the performance of the RSTM output units on the test set (Figure 26).

We presented 10 million noise patterns to all RSTM output units and performed an information theoretic spike triggered average and covariance analysis (iSTAC) (Simoncelli et al., 2004) to estimate the most informative filters. For each filter we also calculated the nonlinearity (Chichilnisky, 2001). Figure 27A shows the estimated LN model for one of the RSTM Spike output units (electrode #12, Monkey M). The spike triggered average (STA) showed a hotspot of sensitivity located in the top left quadrant of the full input space with an increased/decreased activation for a black/white patch as shown with the corresponding nonlinearity (Figure 27C). The similarity between the STA and filter #1 shows that most of the information was carried by the STA (Figure 27D). The next four most informative filters (#2, #3, #4, and #5, Figure 27A) had orientation sensitivity. In contrast to filter #1, the orientation sensitivity was polarity insensitive; stimuli that matched a filter or that matched the polarity inverse of a filter evoked increased responses compared to the mean FR. In addition, filter #4 showed sensitivity for a rapid contrast reversal between -80 ms and -67 ms. Excitatory filters #8, #9, #11, and #12 also showed a contrast invariant excitatory sensitivity for dynamic stimuli with different orientations at different points in time. Finally, multiple filters showed a suppressive or excitatory sensitivity for the lower right quadrant in addition to the top left quadrant. This is best seen in panel B which shows the space-filters (sum over time per filter, see Materials and Methods). The sensitivity filter (sum over all 16 filters and time, see Materials and Methods) in Figure 27E shows two distinct RF location ‘hotspots’ for this unit with strong sensitivity for the top left quadrant and modest sensitivity for the lower right quadrant.

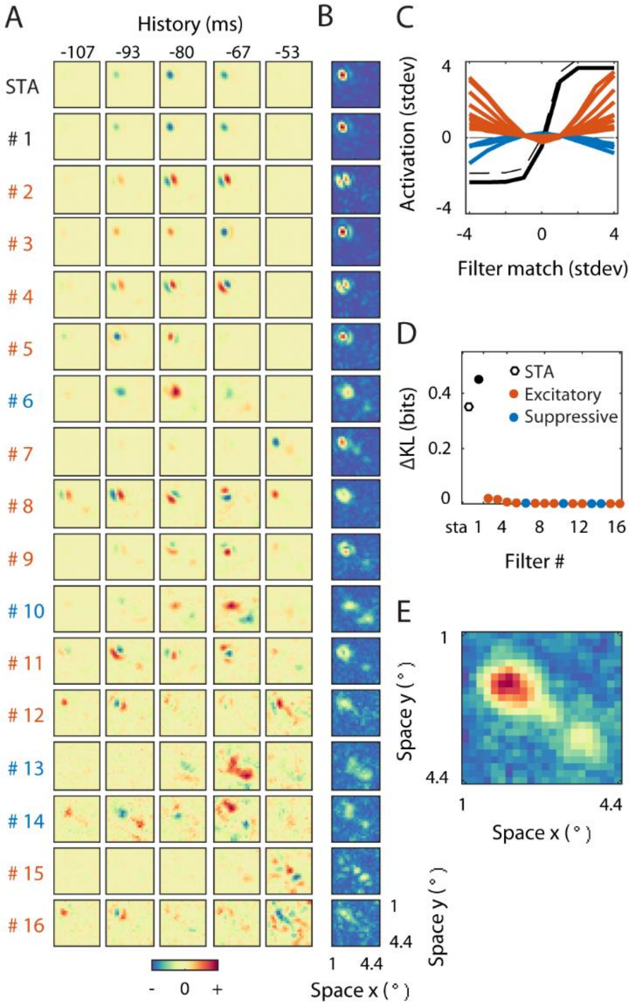

Figure 27. LN model of an example RSTM Spike output unit (electrode #12). A) Linear filters. The spike triggered average (STA) and the 16 linear filters ordered by the amount of information they carry (filter numbers show rank order). Red/blue indicates filters that increase/decrease firing rate above/below the mean activation of the cell. The panels show the space-time filters scaled to the maximum and minimum over time (67 ms) per filter. B) Space-filters, scaled to the maximum and minimum over all space-filters (see Materials and Methods). C) Nonlinearities. The nonlinearities of the STA (black) and 16 most informative excitatory (red) and suppressive (blue) filters. Stimuli that match a filter (high

positive axis projection) or that match the polarity inverse (high negative axis projection) result in a higher FR for the excitatory filters (red) and a lower FR for the suppressive filters (blue). D) Information. The amount of information (in bits) contained in the STA (black), the excitatory filters (red) and suppressive filters (blue). E) Sensitivity filter (sum over all sixteen space-filters, see Materials and Methods). This figure shows that individual RSTM output units captured complex 1st and 2nd order space-time tuning properties of the V1 FMA electrodes; excitation/suppression for multiple dynamic orientations at multiple distinct RF locations (hotspots).

Figure 28 shows the sensitivity filters (see Materials and Methods) for the population of RSTM MUAe output units. Surprisingly, the LN models revealed that most RSTM MUAe output units showed a sensitivity for multiple RF locations (hotspots). This finding was shared by most output units of RSTM Spike and in both monkeys.

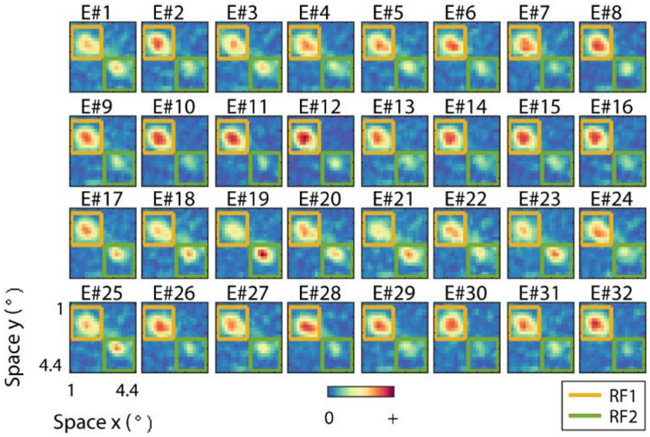

Figure 28. Sensitivity filters of the RSTM MUAe output units. Each panel shows the weighted sum over 16 most informative linear filters and 67 ms (as shown in Figure 27E) for 32 RSTM MUAe output units (Monkey M). Most units showed 1st and 2nd order space-time tuning for two distinct spatial locations (RF1 boxed in yellow, RF2 boxed in green). Some units had strongest sensitivity for RF1 and weaker sensitivity for RF2 (i.e. unit #12), some showed the reverse (i.e. unit #19), or a sensitivity more equally distributed over the two hotspots (i.e. unit #25).

We emphasize that, even though most of the information for the example unit shown in Figure 27 was carried by the STA and was much lower for higher ranked filters, they were all significant due to the estimation with millions of noise patterns. Similarly, the multiple hotspots per output unit shown in Figure 28 reflect real tuning properties. However, with typical measurement noise levels, one could never find such filters in the brain and this does raise the question whether we can show experimentally that the individual and sensitivity model filters are real features of V1 processing.

In addition, the LN models are, by definition, 1st and 2nd order (feedforward) models of the tuning properties for the RSTM output units that might be more complex (i.e. tuning to higher than 2nd order space-time correlations). The LN models are not well suited for further analyses of potential (nonlinear) interactions within and between the observed hotspots. Instead, we used the LN model results primarily as a first estimation of the location of the hotspots for each of the RSTM output units.

In the next sections we show results for more detailed investigation of the hotspots with RSTM simulations aimed to find salient hotspot dependent tuning properties in the orientation, texture, and motion domains. The main focus of these simulations was to estimate the tuning properties for individual hotspots and interaction effects over multiple hotspots. These are shown below in Figure 29, Figure 30, and Figure 31. Next, we tested the RSTM tuning properties with validation experiments where we stimulated V1 three days after the white noise experiment with a selection of the most salient simulated stimuli. The results of the validation experiments are presented in Figure 32, Figure 33, Figure 34, and Figure 35.

4.3.1.2 Orientation tuning of the RSTM output units

The LN models of the RSTM output units revealed complex orientation tuning properties at distinct spatial locations and with variable time course over 67 ms (Figure 27). We investigated the orientation tuning properties in more detail by simulating static oriented noise patterns over five time bins (67 ms). We presented the stimuli to two hotspots (RF1 and RF2) that had the greatest overlap over all RSTM output units (see Figure 28 and Materials and Methods). We estimated the orientation tuning curves for RF1 and RF2 individually, the full 22x22 RSTM input space (RFFull) and RF1 and RF2 combined (RF12) for all orientation conditions.

Figure 29A shows the mean orientation tuning curves (see Materials and Methods) for RF1, RF2, and RFFull of an example RSTM Spike output unit (electrode #12, shown previously in Figure 27). The preferred and non-preferred orientation of this unit when probed on RF1 was 60° and 90°, respectively. When stimulated on RF2 the unit showed modest orientation tuning with a 100° preferred and 150° non-preferred orientation. Stimulated on RFFull the unit had a 80° preferred and 140° non-preferred orientation.

Figure 29B shows the mean orientation tuning interactions over RF1 and RF2 for this example unit (see Materials and Methods). The preferred combination of orientations was 30° (RF1) and 130° (RF2) and the non-preferred combination was 130° and 100° (marked with the plus and open circle, respectively). Note that the maximum activation was higher for orientation combinations (panel B) compared to RF1 and RF2 individually (panel A). This shows that the RSTM output unit was most selective for a specific combination of orientations presented to RF1 and RF2. This interaction effect over RF1 and RF2 was strongest when an 80° angle was presented to RF1 and a 150° angle to RF2 (marked with a cross in panel B).

This shows that the RSTM output unit was selective for specific orientation combinations presented to RF1 and RF2, a tuning property required for the detection of contours (Heydt & Peterhans, 1989; Ito & Komatsu, 2004; Lee & Nguyen, 2001), angles, arcs, and circles (Hegdé & Van Essen, 2004). Interestingly, this is a computation attributed to cortical areas higher up the visual stream such as V2 and V4, not V1. Given that this example RSTM output unit was fitted to a well isolated single V1 neuron, this result raises the question whether the observed interactions are real or an artifact of the RSTM training procedure.

We tested the model predictions for this unit with a V1 validation experiment three days after the white noise experiment (see Materials and Methods). Figure 29C-D shows the orientation stimuli we selected for the validation experiments. We selected the oriented patterns that triggered the maximum (top panels) and minimum (bottom panels) activation at the preferred and non-preferred orientation for RF1, RF2, RFFull, and RF12 (from left to right, RSTM unit activation displayed above and below the stimuli). Figure 29D shows the selected oriented pattern combination that triggered the maximum nonlinear interaction over RF1 and RF2 for the maximum and minimum patterns (the condition marked with a plus in panel B).

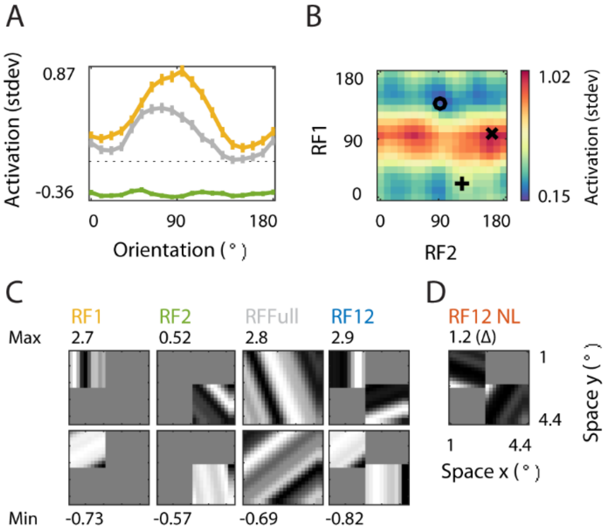

Figure 29. Orientation tuning interactions of an example RSTM Spike output unit. A) Mean orientation tuning curves for hotspot RF1 (yellow), RF2 (green), and RFFull (gray). Error bars indicate standard error over patterns. B) Mean response of the unit to all combinations of orientations presented to RF1 and RF2. The cross and circle mark the

preferred and non-preferred orientation combination. The plus marks the orientation combination with the largest response difference between the linear sum over RF1 and RF2 and RF12. C-D). The patterns that generated the maximum (top row) and minimum (bottom row) response at the preferred and non-preferred orientations for RF1, RF2, RFFull, and RF12 (C), and strongest nonlinear interaction over RF1 and RF2 (D). These stimuli were selected for the V1 validation experiment to test the RSTM predictions for this example unit.

For the population of RSTM MUAe and RSTM Spike output units fitted to the Monkey M and Monkey Y data sets, we found a large variety of orientation tuning properties per RF condition; some units had orientation tuning properties and nonlinear interactions similar to the example unit shown in Figure 29 with much stronger tuning for RF1 than RF2 or the reverse, others had more equal tuning strength for RF1 and RF2. At the end of the results section we will quantify the nonlinear interactions for the population of output units in more detail when we show the comparison between the RSTM predictions and the results of the V1 validation experiments. First, we show the results for a second set of simulations where we tested the tuning interactions over hotspots of the RSTM output units for multi-point correlations.

4.3.1.3 Texture tuning of the RSTM output units

We stimulated the RSTM output units with identical textures as we described previously for RFAM (see chapter 3). We used the five visually salient 2-dimensional texture classes (MSCT) that isolate multipoint correlations previously studied psychophysically

(Hermundstad et al., 2014; Tkacik et al., 2010; Victor & Conte, 1991, 2012). For the random textures, check colors were assigned white or black independently. The *white triangle* and *black triangle* contained three-point correlations and the *even* and *odd* textures four-point correlations (Hermundstad et al., 2014). Similar to the orientation simulations, we investigated interactions over hotspots by presenting the static example textures to RF1, RF2, and RFFull and all texture class combinations to RF12 combined (see Materials and Methods).

Figure 30A shows the mean MSCT selectivity (see Materials and Methods) for RF1, RF2, and RFFull of an example RSTM Spike output unit (electrode #11). When stimulated on RF1, this unit showed a strong selectivity for third order MSCT with a decreased response for *white triangle* textures and an increased response for *black triangle* textures compared to the *random* textures. This unit showed a weaker selectivity for fourth order MSCT with a decreased response for the *even* textures and an increased response for the *odd* textures compared to the *random* textures. We found comparable but reduced and increased MSCT selectivity for third and fourth order textures when we stimulated RF2 and RFFull.

Figure 30B shows the mean MSCT selectivity interactions (see Materials and Methods) over RF1 and RF2 for this example unit. The preferred combination of textures for this unit was *black triangle* textures on RF1 and RF2 and the non-preferred combination was *white triangle* textures on RF1 and RF2 compared to *random* textures (marked with a cross and open circle, respectively). The maximum nonlinear interaction over RF1 and RF2 for this unit was when a *white triangle* was presented to RF1 and a *black triangle* texture to RF2 (marked with the plus symbol). Note that the relatively low response range shown in panel A and B is based on the mean over 5000 examples per texture class and 67 ms. This weak modulation is most likely due to selectivity for only a small subset of examples per texture class. In other words, the unit does not respond strongly to every example of the black triangle class. It would be challenging to demonstrate such MSCT selectivity using a random selection of examples per texture class. The advantage of our ANN approach, however, is that it allows us to extract the most salient examples of the preferred and non-preferred texture classes of the RSTM model units. Figure 30C-D shows the most salient examples per RF condition that we used for the V1 validation experiment discussed below (same convention as Figure 29C-D).

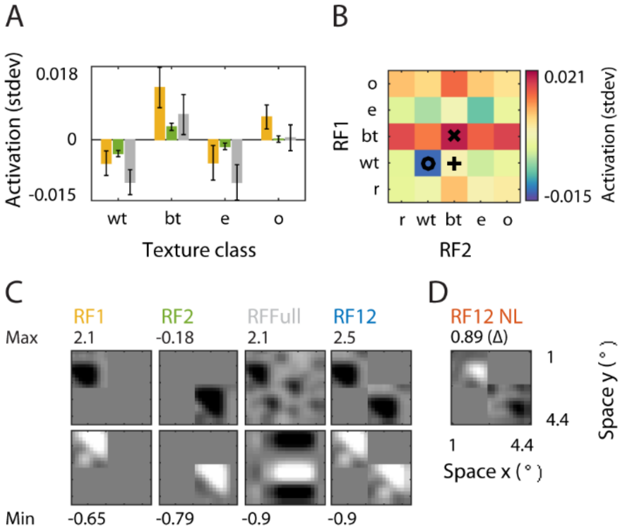

Figure 30. Texture tuning interactions of an example RSTM output unit. A) The mean MSCT selectivity to four multipoint textures (white triangle, black triangle, even, and odd)

is shown with the response to random textures subtracted for RF1 (yellow), RF2 (green), and RFFull (gray). Error bars indicate standard error over patterns. B) Response of the unit to all combinations of textures presented to RF1 and RF2. The cross and circle mark the preferred and non-preferred texture combination. The plus marks the texture combination with the largest response difference between the linear sum over RF1 and RF2 and RF12 (i.e. strongest nonlinear interaction over RF1 and RF2). C-D). The patterns that generated the maximum (top row) and minimum (bottom row) response at the preferred and non-preferred texture classes for RF1, RF2, RFFull, and RF12 (C), and the strongest nonlinear interaction over RF1 and RF2 (D). These stimuli were selected for V1 validation to test the RSTM predictions for this example unit.

We found a large variety of MSCT selectivity for the RF conditions; some units had MSCT selectivity comparable to the example unit shown in Figure 30 with a stronger selectivity for RF1 than RF2 or the reverse while others had more equal selectivity strength for RF1 and RF2. At the end of the results section we will quantify the nonlinear interactions for the population of output units in more detail when we show the comparison between the RSTM predictions and the results of the V1 validation experiments. First, we show the last set of model simulations where we tested the motion tuning interactions over hotspots of the RSTM output units.

4.3.1.4 Motion tuning of the RSTM output units

We investigated hotspot interactions of the RSTM output units with motion stimuli for all RF conditions. Figure 31A-B shows the mean motion tuning (see Materials and Methods) for RF1 and RF2 for an example RSTM MUAe output unit (electrode #23). When stimulated on RF1, this unit showed a low pass motion tuned response with a preferred speed of 3.75°/s in downward direction (90°). The non-preferred speed and direction was 30°/s in leftward-upward direction. The motion tuning properties were different for RF2; band-pass speed tuned with a preferred speed of 30°/s in rightward direction and a non-preferred speed of 30°/s in leftward direction.

Figure 31C shows motion tuning interactions over RF1 and RF2 for this example output unit. The figure shows how all motion conditions presented to RF1 and RF2 interact. For this unit, the combined preferred speed and direction tuning was comparable to the motion tuning properties for individual hotspots. However, the maximum response over motion conditions presented to both hotspots was more than twice the maximum response over motion conditions presented to individual hotspots. In other words, if RF1 and RF2 were independent, the speed and motion direction presented to RF1 (x-axis Figure 31) would not affect the response to the preferred speed and direction presented to RF2 (y-axis Figure 31). The more than two-fold increase at the preferred speed and direction for RF2 (compare maximum activation panel B and panel C) suggests a highly nonlinear interaction over hotspots in the motion domain.

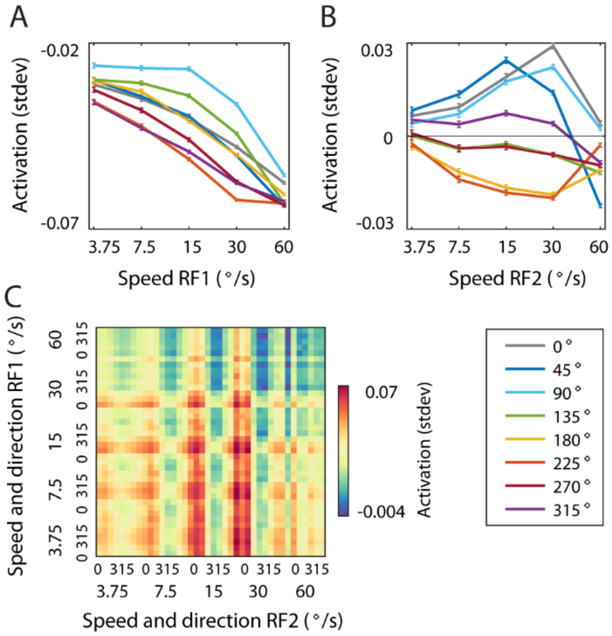

Figure 31. Motion tuning interactions of an example RSTM MUAe output unit. A-B) The mean motion tuning (speeds 3.75, 7.5, 15, 30, and 60°/s and directions 0°, 45°, 90°, 135°, 180°, 225°, 270°, 315° where 0° is defined as rightward motion) is shown for RF1 (A) and RF2 (B). Error bars indicate standard error over patterns. C) Mean motion interaction effects for RF1 (y-axis) and RF2 (x-axis) over all motion condition combinations. This figure shows that speed tuning and direction selectivity for RF1 and RF2 is stronger when a particular speed and direction combination is presented to RF1 and RF2.

For the population of RSTM MUAe and RSTM Spike output units fitted to the Monkey M and Monkey Y data sets, we found a large variety of motion tuning properties for the RF conditions; some units had motion tuning properties comparable to the example unit shown in Figure 31 with strong motion tuning and nonlinear motion interaction effects over RF1 and RF2 while others showed motion tuning for RF1 or RF2 individually, but not both.

In the next few sections we will describe the orientation, texture, and motion interaction effects in more detail when we show the experimental validation of the RSTM predictions in V1.

4.3.2 Experimental validation of the RSTM predictions in V1

4.3.2.1 RSTM-V1 comparison for an example unit with optimized stimuli

Based on model simulations we selected the most salient orientation, texture, and motion stimuli and presented them to V1 (see Materials and Methods). Figure 32 compares the time course of an example RSTM MUAe output unit (electrode #28, panels on the right) and corresponding V1 response (left panels) in response to the stimuli that were predicted to maximally or minimally drive this unit when presented to individual and hotspot combinations (see legend).

We define a similar performance measure as the train and test set of the experimental noise stimuli (see Figure 26) for the RSTM predictive performance based on the explained variance between the response of the RSTM output units and the response of the V1 validation experiment as a fraction of the V1 consistency (see Materials and Methods). For the example RSTM output unit shown in Figure 32, the RSTM predicted time course matched the V1 response well for most stimulus and RF conditions. In other words, the variance of the V1 response that was captured by the RSTM output unit was close to the explained variance of individual V1 responses compared to the other responses to identical stimuli/repeats (R^2=73%, R^2=65%, and R^2=85% for all orientation, texture, and motion stimuli, respectively).

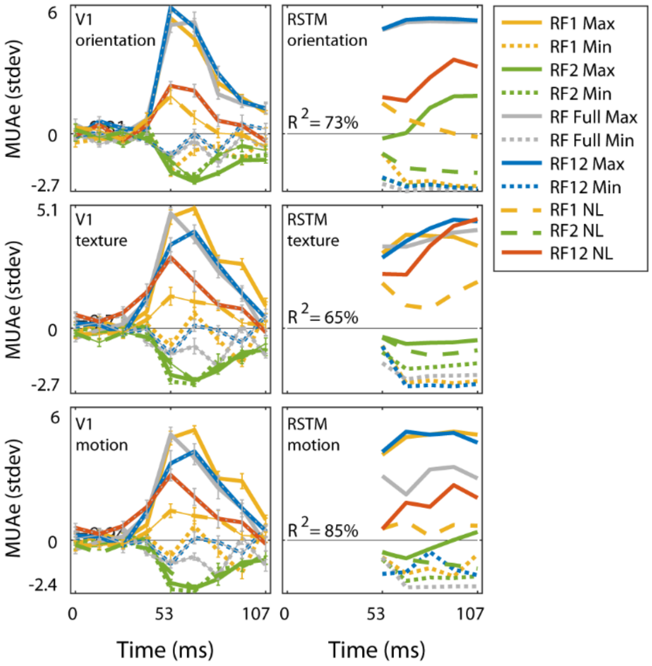

Figure 32. RSTM and V1 experimental validation of an example electrode. Response over time (MUAe electrode #28) of V1 (left panels) and corresponding RSTM MUAe output unit to the selected orientation (top), texture (middle), and motion (bottom) stimuli over time (stimulus onset is at 0 ms and average V1 latency is 53 ms). The time course is shown for a presentation of the maximum (solid) and minimum (dashed) predicted stimuli presented to RF1, RF2, RFFull, and RF12, and for the maximum predicted nonlinear interaction over RF1 and RF2 (RF12 NL). For this unit and stimuli, the RSTM unit predicted the V1 response well, including the time course.

The 15 RSTM MUAe units that we used to selectively target individual V1 neurons with the most salient stimuli predicted more than 60% of the variance of the V1 validation responses (as a fraction of consistency, see Materials and Methods, R^2=48%, 73%, and 67% for orientation, texture, and motion stimuli, respectively). Interestingly, some of the 15 RSTM Spike units predicted the V1 response to orientation, texture, and motion stimuli better than the consistencies of the V1 response (R^2=94%, 109%, and 105%, for orientation, texture, and motion stimuli, respectively). This is most likely due to the non-Gaussian (Poisson)-like spiking distributions that are compared to the noiseless smooth output activation values of the RSTM that was fitted to a large stimulus set for that neuron (Discussion).

This result shows that a significant fraction of the RSTM predicted complex properties in the orientation, texture, and motion domains are real V1 tuning properties. As we mentioned previously, the 32 electrodes of the array recorded the responses of all units/neurons and for all (990) experimental stimuli that we presented during the validation experiment. In the last sections we will cover comparisons between the RSTM and V1 with much larger data sets than we have shown thus far. For these analyses we include the experimental stimuli and responses that were originally selected for other units which challenges the RSTM generalization performance far beyond analyses thus far.

4.3.2.2 RSTM-V1 comparison for an example unit for all stimuli

Figure 33A shows the explained variance (as a fraction of the V1 consistency for these stimulus sets, see Materials and Methods) between the model predictions and V1 for an example RSTM MUAe output unit (electrode #11). The figure shows the explained variance for the 30 maximum (first bars) and minimum (second bars) responses when stimuli were presented to RF1 (yellow), RF2 (green), RFFull (gray) and RF12 (blue).

Figure 33B shows the data points that were used for the R^2 values in panel A. The example RSTM unit predicted the V1 response well for a large fraction of the stimulus set and for all stimulus and RF conditions, in particular the texture stimuli (average R^2=80% over the eight RF conditions). The response to eight stimuli that were specifically selected for this unit are marked with open circles (minimum) and diamonds (maximum).

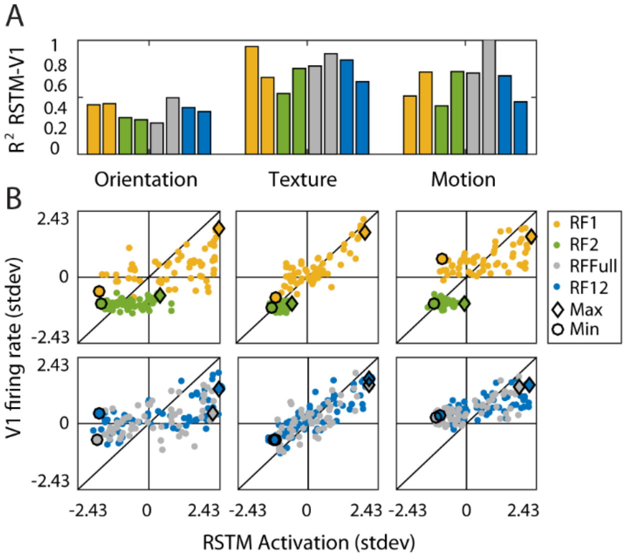

Figure 33. RSTM and V1 validation experiment of an example unit. A) The explained variance between the RSTM predictions and the V1 response to the 30 selected minimum (first bars) and 30 selected maximum (second bars) orientation (left), texture (middle), and motion (right) stimuli presented to RF1 (yellow), RF2 (green), and RFFull (gray) for the RSTM Spike unit shown in Figure 29 (electrode #12). This figure shows that the RSTM predictions generalized well, even to stimuli that were not specifically selected to maximally suppress or excite this unit. B) Scatter plots for the data points in A with the maximum and minimum orientation, texture, and motion stimuli for this example unit marked with diamonds and open circles, respectively.

On a population level (the 11 high performing RSTM units, see Figure 26) the RSTM MUAe explained 41%, 59%, and 51% of the variance of the V1 validation responses (orientation, texture, and motion stimuli, respectively). For the RSTM spike (same units) the RSTM explained 45%, 54%, and 49% for the three stimulus conditions. For monkey Y we found comparable explained variances between the population of RSTM Spike and RSTM MUAe output units and the V1 validation responses (R^2 RSTM Spike = 53%, 55%, and 62%, R^2 RSTM MUAe = 49%, 42%, and 54%).

4.3.3 Testing the RSTM predictions for tuning interactions over hotspots.

The RSTM output units revealed nonlinear response dynamics when specific stimulus combinations were presented to multiple hotspots that could not be explained with a simple summation over the responses when stimuli were presented to individual hotspots (see

Figure 29, Figure 30, and Figure 31). In a more formal way, the response to RF1 + RF2 was not equal to the response to RF12. We tested the nonlinear interactions with a validation experiment in V1. Figure 34A shows the explained variance between an example RSTM MUAe output unit (electrode #7) and V1 validation response to all orientation, texture, and motion stimuli that were selected for a collection of RSTM units and V1 neurons that proved particularly useful when exploring the nonlinear weak interaction effects (see Materials and Methods).

Figure 34B shows the data points that were included in the R^2 values in panel A where we plotted the response to RF1, RF2, and RF12 as a function of RF1 + RF2 to highlight the nonlinearity of the responses. In other words, if the response to RF12 (the red dots in panel B) match the linear sum of the response to RF1 + RF2, the data points would lie on the diagonal. The marked data points indicate the maximum nonlinear response between RF1 (open circle) + RF2 (square) and RF12 (diamond) triggered by the stimulus selected specifically for this unit. In sum, this figure shows that V1 neurons are tuned to multiple hotspots with nonlinear interactions effects, closely following the predictions of the RSTM output units.

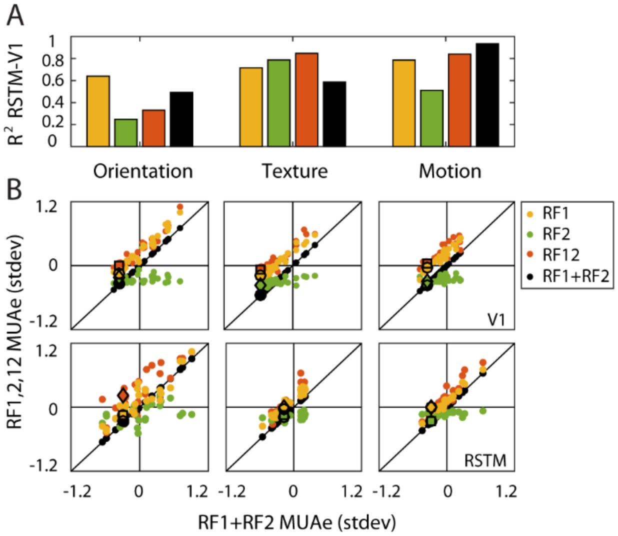

Figure 34. RSTM and V1 validated interaction effects of an example unit. A) The explained variance between the RSTM predictions and the V1 response to the 30 maximum interacting orientation (left), texture (middle), and motion (right) patterns presented to RF1 (yellow), RF2 (green), RF12 (red), and the sum over RF1 and RF2 (black). For this output unit (electrode #7) the RSTM predicted the V1 response well for all stimulus conditions, in particular the texture and motion interactions. B) Scatter plots for V1 (top) and RSTM (bottom) of the response to RF1, RF2, and RF12 plotted against the sum over RF1 and RF2. This figure shows that RF1 and RF2 interact nonlinearly for all tested stimulus conditions (i.e. the RF12 data points are not on the diagonal, which represents the linear sum over RF1 and RF2).

Finally, we tested whether the nonlinear interaction effects over hotspots is also a V1 population (average) tuning property (based on the 11 high performing RSTM Spike output units, see Figure 26). Figure 35A shows the explained variance between the RSTM and the

V1 population average for 30 stimuli presented to RF1, RF2, RF12, and RF1 + RF2 (correlation values for individual units are shown in Figure 35B). The figure shows that the RSTM population average predicted the V1 population average well, especially for the texture stimuli (middle panel). Figure 35C shows the data points that were included in panel A for V1 (top) and the RSTM (bottom) for RF1, RF2, and RF12 plotted against RF1 + RF2 (same convention as Figure 34B). For both V1 and the RSTM, the response to RF12 was much higher than the response to RF1 + the response to RF2. This result shows that the nonlinear interaction effect over hotspots, predicted by individual RSTM output units and confirmed in V1 with validation experiments, was much stronger for the population average (for both the RSTM and V1, compare the distance between all the red RF12 data points and the black RF1 + RF2 data points for Figure 34 and Figure 35).

We found comparable nonlinear interaction effects for the population average of the RSTM MUAe output units and explained variance with the V1 validated responses (R^2 = 22%, 77%, and 46% for orientation, texture, and motion, mean over the four RF conditions). For Monkey Y we found comparable nonlinear interaction effects for individual units and the population average of RSTM Spike and RSTM MUAe, albeit with a reduced explained variance between model and V1 (R^2 = 38%, 34%, and 60% for orientation, texture, and motion, mean over the four RF conditions for RSTM Spike, R^2 = 27%, 41%, and 36% for orientation, texture, and motion, mean over the four RF conditions for RSTM MUAe).

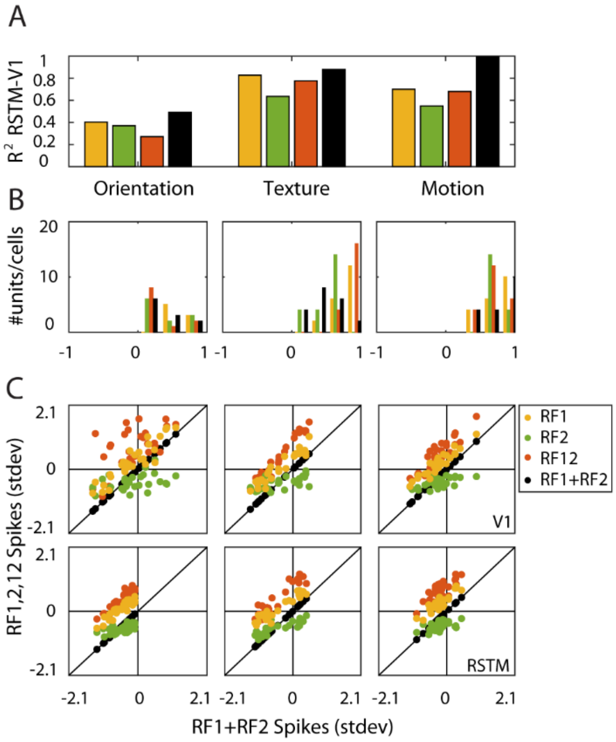

Figure 35. RSTM and V1 validated interaction effects of the population average. A) The explained variance between the RSTM predictions and the V1 response to 30 maximum interacting orientation (left), texture (middle), and motion (right) patterns presented to RF1 (yellow), RF2 (green), RF12 (red), and the sum over RF1 and RF2 (black) averaged over 11 high performing RSTM Spike output units and corresponding isolated V1 single cells. The RSTM predicted the V1 response well for all stimulus conditions and RF locations, especially for the texture stimuli. B) Histograms of the correlations that were used in the panels in A for all RF and stimulus conditions. C) Scatter plots for the data points in panel A. This figure shows that the RSTM predicted and V1 confirmed nonlinear interactions over RF locations is stronger for the population average than for individual units.

4.4 DISCUSSION

We investigated V1 space-time tuning properties with our ANN approach. First, we presented dynamic white noise stimuli to V1 of the awake behaving macaque that allowed us to extract the V1 response properties in a data-driven way. Next, we fitted an artificial recurrent neural network (RSTM) to the experimental noise stimuli and V1 responses. Third, analyzing the RSTM with simulations of several stimulus classes revealed that the output units were selective for multiple RF locations (hotspots) and with nonlinear interactions in the orientation, texture, and motion domains. Finally, we experimentally validated the model predictions with the most salient stimuli for the high performing RSTM units that confirmed the existence of and nonlinear interactions over multiple hotspots in V1.

In the next sections we will discuss the performance of the RSTM on the train and test set of the experimental white noise and V1 validation stimuli, the nonlinear interaction effects over multiple RF hotspots, and the functional extend of the RSTM recurrent connections.

4.4.1 Performance of the RSTM

The main motivation for the experimental noise stimuli was to explore the V1 tuning properties in a data-driven way. In theory, all V1 tuning properties bounded by the luminance range and spatiotemporal frequency of the noise stimuli (simple form, complex form, and motion) can be captured. One significant disadvantage of this approach, however, is to determine whether the properties of the fitted model are real, noise within the data set, or an artifact of the fitting process (i.e. local minima). We solved this problem with the well-established approach by fitting the RSTM to a train set (90%) and use the test set (10%) primarily to keep track of the generalization performance to out-of-sample noise patterns. By definition, this objective performance level also defines a level of confidence for generalization to all patterns within the spatiotemporal frequency bounds of the noise stimulus set, including the stimuli that were used for the V1 validation experiments. It is, therefore, not surprising that the high performing units on the test set generalized well to the validation stimuli.

We defined the performance on the train and the test set as a ratio of consistency, the amount of variance explained by a single V1 response over all other responses to identical noise stimuli (see Materials and Methods). In other words, consistency defines an upper bound for the amount of variance in the V1 response that can be captured by the model. Interestingly, some RSTM MUAe and many RSTM Spike output units had a higher performance than the V1 consistency. This suggests that for those units, the RSTM mapped the noise input pattern streams with a nonlinear fit over V1 repeats. In other words, the noiseless smooth output activation values of the RSTM that was fitted to a large stimulus set for that neuron, provided a better generalization performance than the non-Gaussian (Poisson)-like spiking response distributions.

In contrast, some RSTM output units had a reduced performance on the train and test set as a ratio of the V1 consistency; for these units, the RSTM did not capture all the variance of the V1 responses. Because the higher V1 consistency suggests that the low performance is not due to external (measurement) noise, we believe that there is significant room for improvement of the model fit. For instance, by using a different architecture with more neurons or additional layers, or an improved training algorithm.

4.4.2 V1 validation experiments

The four step ANN approach allowed us to extract the most salient examples of the preferred and non-preferred orientation, texture, and motion stimuli for the RSTM units (Figure 29, Figure 30, and Figure 31, respectively) that we validated with V1 experiments. The high predictive performance of the RSTM units for the V1 responses was not surprising, given the high performance on the test set of the experimental noise pattern streams. This result does, however, show that the RSTM generalized to patterns far removed from the experimental noise sample in a robust manner.

The selection of most salient stimuli in the orientation, texture, and motion domain were, by definition, the stimuli that maximally drove (excite or suppress) the responses of the RSTM and V1 units. This specific stimulus set explains the performance discrepancy between the RSTM train and test set (Figure 26) and the RSTM and V1 validation responses (Figure 32 and Figure 33). Whereas the former defines the performance for all possible inputs, including the patterns that the units don’t care about, the latter defines the performance for patterns that the units maximally cares about. Interestingly, this performance difference was highest for the RSTM Spike output units (sometimes three-fold the V1 consistency, data not shown). This strongly suggests that our ANN approach can bring to light highly specific cell tuning properties that are challenging to (consistently) measure directly.

Most RSTM units were sensitive to multiple distinct RF hotspots. However, this RSTM property had a reduced predictive power for the V1 validation experiments (Figure 32, Figure 33, and Figure 34). We believe this is most likely due to a much lower response modulation and weak interaction effects for the second compared to the first hotspot (i.e. a lower signal to noise). We emphasize, however, that the high explained variance for RF12 is sufficient to show the main finding that V1 computes more complex shapes than current knowledge about the classical RF suggests (Figure 34 and Figure 35). We will discuss this topic in more detail in the next section.

4.4.3 Nonlinear interactions over hotspots

The RSTM predicted that the summed response for orientation, texture, and motion stimuli presented to RF1 and RF2 was different from the combined presentation to RF12. In other words, the RSTM predicted nonlinear interaction effects for both the individual units and for the population average over multiple distinct hotspots (Figure 34 and Figure 35). With validation experiments we confirmed these complex V1 tuning properties. What could be the functional role for sensitivities to higher order spatial and space-time correlations over multiple distinct RF locations?

Given that the interaction effects were relatively weak, it is not surprising that little is known about the existence and role of these interactions. Interestingly, our findings do show parallels with the work of (Das & Gilbert, 1999) who investigated ‘contextual modulations’ mediated by short-range interactions between neighboring V1 neurons. They hypothesized that the location within the cortical map of orientation and space would largely influence local visual processing, in particular computations required for complex form detection. The large degree of variability in homogeneity of the orientation maps, from relatively constant to regions with big orientation shifts such as reversals, sharp jumps (fractures) and point singularities (Blasdel & Salama, 1986; Bonhoeffer & Grinvald, 1991; Hubel & Wiesel, 1974; Ts’o et al., 1986), should be reflected in local V1 function. Combined with anatomical studies that showed that the dendritic and local axonal arborizations of pyramidal neurons extend laterally out beyond the presumed functional boundaries such as orientation singularities (Malach, Amir, Harel, & Grinvald, 1993) or other features within V1 (Hubener & Bolz, 1992; Katz, Gilbert, & Wiesel, 1989; Malach, 1992), they hypothesized that neurons close to orientation shifts are likely to be functionally connected to neurons preferring a range of orientations, including orthogonal ones. With optical imaging in cat V1, they found that the pattern of local connectivity was indeed related to contextual modulations; V1 neurons were sensitive to lines orthogonal to their preferred orientation within their classical receptive field. We confirm and build upon their findings with our ANN approach applied to monkey V1 in which we provide a more detailed description of the kind of visual stimuli and network function in the orientation, texture, and motion domains.

4.4.4 Effective memory of the RSTM recurrent connections

In the previous sections we discussed the role of the nonlinear interactions over multiple distinct hotspots for single RSTM output units and V1 neurons. An important question, one that can be asked for the RMM (Motion detection based on recurrent network dynamics) and the RFAM (chapter 3) as well, is about the source of the underlying nonlinear interactions over time and the relationship with the neural systems. For the ARNNs, the answer is straightforward; by design, any response interaction over time must be rooted in the lateral/recurrent connectivity of the hidden units. In other words, a complete removal of the lateral/recurrent connections would automatically result in linear response properties for the hidden and output units; a static response when a static input is presented or a dynamic response when a dynamic input is presented, but equal to the sum of the responses to the components of the input.

To quantify the nonlinear contribution of the recurrent connectivity (both in strength and in duration), we ran 1000 simulations of two consecutive white noise pulses that covered the full input space and were presented with variable temporal delays (Figure 36). Panel A shows the mean response over the 1000 two-pulse white noise inputs with the response to the first pulse subtracted from the response to the second pulse and up to nine time bins

(13-120 ms) temporal delays (x-axis) for an example RSTM MUAe output unit (electrode #30). The figure shows that the recurrent connections for this unit had a nonlinear effect up to approximately eight time bins (107 ms) in between the first and second pulse; a slight suppression for a 13 ms delay and strong excitation from 26 ms onward with a maximum at a 40 ms delay between the first and the second pulse. Figure 36B shows comparable nonlinear interaction effects for the population of high performing RSTM MUAe output units.

This result is significant because the two-pulse interaction plots show that the RSTM computes input pattern streams in a complex nonlinear way and over multiple time windows (up to around eight time bins). These higher order interactions cannot be captured by feedforward models that only take into account two-point space-time interactions such as (any version) of the ME model (Adelson & Bergen, 1985; Watson & Ahumada, 1985). This further highlights the previously mentioned limitations of the 1st and 2nd order based LN-models (Figure 27). Of course, to fully capture the response dynamics, the 1st and 2nd order LN models could be extended by explicitly including up to 8th order space-time interactions. However, the curse of dimensionality would quickly enforce practical limitations beyond 2nd order, even for modern day hardware.

The close match between the response of the RSTM output units and the V1 neurons for both the white noise and the V1 validation stimuli suggests a similar match for the two-pulse interaction simulation (Figure 36). If the RSTM output units generalize well to two-pulse interactions, similar conclusions can be drawn for V1. And, of course, these higher order space-time tuning properties can always be described by feedforward models. However, we showed that lateral and recurrent connections could be a more efficient source for spatiotemporal integration, especially when taking into account the ubiquitous cortical connectivity and the evolutionary pressure to compute with the least amount of energy and neural hardware.

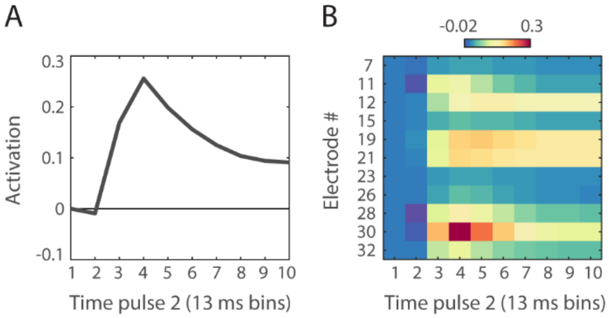

Figure 36. Two pulse interaction effects over variable time intervals. A) Response of an example RSTM output unit (unit #30) to two white noise pulses with intervals up to nine time bins. The first pulse was always presented at time bin one and the second pulse at time bins two-ten (x-axis), gray values otherwise. The figure shows the response to the second pulse with the first pulse subtracted. For this example unit, the first pulse had a nonlinear interaction effect on the second pulse with intervals up to around eight time bins. B) Response of eleven high performing RSTM output units (y-axis) to the second pulse at one of ten intervals (x-axis) with the response to the first pulse subtracted. The figure shows that, for most RSTM output units, the first pulse had a nonlinear effect on the second pulse up to an interval of eight time bins.

5 GENERAL DISCUSSION

5.1 DATA-DRIVEN ANN APPROACH

In this thesis we developed an approach to do data mining on experimental neural data using our ANN approach. As a proof of concept we fitted an ARNN – the RMM - to motion tuned response properties of macaque area MT neurons (chapter 2). This modeling attempt provided the novel insight that recurrent connectivity can explain the sequential recruitment phenomenon (Mikami, 1992) found in MT neurons, without the need for feedforward temporal delay lines. However, this study also brought to light limitations of the MT data set. In particular, the large RF size of MT cells forced us to reduce the motion input stimuli to one dimension. Such adjustments force a kind of assumption that we wish to prevent with our data-driven ANN approach.

In chapter 3 we applied the ANN approach to V2 neurons with a selectivity for multipoint correlated textures (MSCT). With the much smaller RF size of V2 neurons we could overcome one of the MT data set limitations by fitting the ARNN – the RFAM - to each of the experimental stimuli. Unfortunately, noisy measurements did not allow us to map individual experimental stimuli onto individual V2 responses. Instead, we were forced to model the mean V2 response to all stimuli instead, assuming a homogeneous response to all experimental stimuli. we found that the RFAM hidden units had complex MSCT selectivity properties that was correlated with motion tuning strength. The overlap between form and motion processing was a robust generalization of the RFAM, suggesting that tuning to higher order spatial and space-time correlations could also be shared by real neurons in the visual system. This prediction awaits experimental confirmation and brings to light a more pressing limitation of the first two data sets; they were collected with recordings where the neurons were lost after each session and could not be used to validate model predictions.

For the third project we applied our ANN approach to chronically implanted microelectrode arrays (FMA) in V1 of the awake behaving macaque. With the FMAs we were able to apply a data-driven approach without the need to assume the previously discussed biases. We collected experimental data with high resolution dynamic noise stimuli, fitted an ARNN - the RSTM - to the exact experimental inputs and outputs, extracted the tuning properties of the fitted model units, and test (the most salient) stimuli with a validation experiment on the same neurons we used to fit the model a few days after the white noise experiment. With this approach we revealed that complex RF tuning properties arise in V1 with nonlinear interactions over multiple distinct hotspots. This strongly suggests that V1 computes complex shapes such as corners, arcs, and contours (i.e. higher order spatial correlations) that is currently primarily attributed to higher cortical areas in the visual processing pathway such as V2 and V4 (Freeman et al., 2013; Qiu & Heydt, 2005).

5.2 ANN AND RULE-BASED MODEL APPROACH

The main advantage of the rule-based approach is that the designer has a relatively complete understanding of the basic properties that underlie the model. If the number of free parameters (or rules) is small, predictions of the model can often be traced back to one or more of those basic properties. The disadvantage of this approach is that only regularities that can easily be phrased as ‘rules’ are implemented in the model, resulting in strong biases towards oversimplification and confirmation of interpretations, rather than to challenge them.

The fundamentally different ANN approach we proposed starts with actual data, not the “rule” that the experimentalist has used to describe or summarize the data. In contrast to the rule-based models, the solution of the ANN is typically surprising and less intuitive. We showed, however, that a detailed investigation of the network elements can reveal novel insights for the already well understood neural response properties and complex dynamics that are harder to capture with simple rules (i.e. sequential recruitment beyond the first two motion steps (McKee & Welch, 1985; Mikami, 1992)).

We see significant room for improvement when the strengths of the rule-based approach are combined with the basic ANN design we used in the current thesis. For instance, the network designs we discussed had all to all connectivity between many units that processed their combined input in an identical way. These properties are far removed from known connectivity patterns and the large diversity of neurons found in the real brain that have been implemented as additional rules to the ANNs we used for our approach. In the remainder of this section we will highlight some of these ‘hybrid’ model designs and propose how they can be used to advance the field of systems neuroscience.

5.3 INTRINSIC TEMPORAL DELAYS

One of the key components of a motion detector is a mechanism for keeping track of the time it takes for an object to travel from point A to point B (the effective dt). In the ME model (Adelson & Bergen, 1985; Watson & Ahumada, 1985) this is implemented with feedforward temporal delay lines and in the motion model of (Maex & Orban, 1996) with intrinsic cell properties that vary in their temporal response dynamics. In the RMM, we investigated whether the ubiquitous recurrent connections in cortex could play the critical role of temporal delays for motion detection (chapter 2). Therefore, we modeled our measured MT data with an ARNN with identical units that process the motion signal within the same integration time window (the frame rate of the monitor that was used during the experiments). The recurrent connections of the network allowed the model to self-organize temporal delays such that the MT responses to various speeds and motion directions could be captured.

In a biological neural network these effective delays could interact with the intrinsic delays of subpopulations of neurons or slow excitatory and fast inhibitory synaptic transmission (the source of temporal delays in the motion model of (Maex & Orban, 1996)). The simple ARNN we used did not allow for an investigation of such temporal interactions but this limitation can be overcome with a different ANN design such as the feedforward time-delay network (Lang, Waibel, & Hinton, 1990). Here, temporal delays can be hardcoded in individual network components. Because the ARNN can generate delays quite flexibly it is not clear whether this would add computational power to the network, but it certainly adds biological realism and may lead to novel experimental predictions.

5.4 NETWORK WEIGHT RESTRICTIONS

A second network property that can be matched better to known properties of the neural system is the weight distribution of the connections between the network units. For our motion model (chapter 2) the units were all to all connected with randomly initialized weights. The main benefits of a fully connected network with random initial weights are practical; in our experience the network converges to a solution more rapidly. Because there are many ways to generate a specific input-output relationship, an unrestricted network will most likely converge to a different solution than the neural system it aims to model. This can nevertheless be useful to gain insight into possible solutions, but it may be beneficial to guide the network to a more biologically plausible solution by restricting network weight parameters based on known properties of the neural system.

For instance, the output weights of each unit of the RMM, RFAM, and RSTM had both positive and negative weights. This conflicts with the strictly excitatory or inhibitory nature of neurons in the brain and prevented us to infer any specific role of the two classes of neurons for visual processing that might have been possible with a restriction to strictly positive or strictly negative weights for separate subsets of network units.

Additional weight restrictions might also be more suited for a simulation of the hierarchical organization of object recognition of the visual system. Along the ventral pathway, neurons in early visual areas (LGN-V1) are more sensitive to lower-order visual features such as luminance edges while neurons in later areas (V2-V4) show selectivity for higher-order features such as corners, curvatures and eventually complex object shapes (Anzai et al., 2007; De Weerd, Desimone, & Ungerleider, 2003; Hegdé & Van Essen, 2000; Hegdé & Van Essen, 2004). To model the hierarchy, a class of ‘deep-learning’ ANNs with strong weight restrictions can be more suited than a fully connected neural network such as the ARNNs we used. Inspired by systems neuroscience, deep neural networks have gained popularity for their superior performance in classification tasks such as visual object recognition. In feedforward convolution networks (LeCun et al., 1998) such hierarchical organization is implemented with weight restrictions enforcing local connectivity between neurons of adjacent layers (similar to a convolution of the inputs as we previously discussed).

Despite the proven classification power of convolution networks, however, we believe that a strictly feedforward network cannot be a biological plausible model for neural systems that are known to be dominated by recurrent network dynamics. One might argue that any recurrent network based response dynamics, can, in principle be captured by a feedforward (deep layered) network as has been formalized in detail by (Liao & Poggio, 2016). More appropriate for our ANN research approach in systems neuroscience could be hybrid networks such as (Liang & Hu, 2015) where the weight restrictions of the feedforward convolution neural network are combined with recurrent connections within each layer of the network.

5.5 RECURRENT FEEDBACK FROM HIGHER CORTICAL AREAS

In addition to the lateral and recurrent connections within cortical brain areas, neurons in lower cortical areas are also influenced by recurrent feedback from higher cortical areas. For instance, the V1, V2, and MT neurons not only receive feedforward input from the LGN and V1, they are also heavily modulated by recurrent input from neurons within higher cortical areas such as MST (Maunsell & van Essen, 1983). The ARNN designs we used to investigate the three data sets did not have recurrent feedback connectivity (i.e. from output units to hidden units) but this limitation can be overcome with a network design that allows for bidirectional connections between units (Schuster & Paliwal, 1997).

In addition, theoretical and empirical evidence has shown that a single set of recurrent weights in basic recurrent neural networks are suffering from a vanishing gradient that poses strict limitations to the maximum length of pattern sequences that the network can learn (Bengio, Simard, & Frasconi, 1994). In other words, the memory length of the network is limited by the number of recurrent weights. For the short time window of interest that we investigated with our ARNN models (i.e. five simulation time steps for the

RMM and RFAM and eighteen for the RSTM), one set of recurrent weights proved sufficient.

More complex data sets with both short term effects (i.e. feedforward and recurrent connections within MT, V2, and V1) and long term effects (i.e. recurrent feedback from higher cortical areas) would require additional network components. The class of long short-term memory (LSTM) networks were designed to overcome the problem of vanishing gradients (Bowman, Manning, & Potts, 2015; Hochreiter & Schmidhuber, 1997; Kalchbrenner, Danihelka, & Graves, 2015). The LSTM network consists of hidden units (called memory cells) that can remember their inputs for an unlimited time. The memory cell behaves like a gated leaky neuron; it has a recurrent connection with weight one that is multiplicatively gated by another unit that can clear its content at any time. We believe that LSTM networks combined with bidirectional connections could be a powerful research tool to investigate the role of short term and long term recurrent network dynamics. For instance, one theory is that recurrent feedback may perform hypothesis-testing by sending Bayesian priors from higher up the visual hierarchy to lower visual areas (Lee & Mumford, 2003). Although elegant and biological conceivable, it is not clear how predictions and errors are implemented by neurons and how the more abstract language of higher cortical areas translates to the more detailed representations of lower cortical areas. Here, our ANN approach might provide a better view on a potential role for recurrent feedback.

5.6 CONCLUSION

The wider significance of our ANN approach is that it demonstrates that ARNNs can provide a more complete understanding of how information is processed in the brain, and how complex computations emerge from recurrent neural hardware. Although the focus of the thesis was on the neural mechanisms of early visual form processing, the approach has wide applicability across many domains of sensory processing and possible to higher cognitive function as well. The main limitation that the network properties of our basic ARNN are far removed from the real neural system can be overcome with more recent ANN designs that, inspired by advances in neuroscience, rapidly bridge the distance between artificial and biological neural system.

6 REFERENCES

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, *2*(2), 284–299.

Albright, T. D. (1984). Direction and orientation selectivity of neurons in visual area MT of the macaque. *Journal of Neurophysiology*, *52*(6), 1106.

Anzai, A., Peng, X., & Van Essen, D. C. (2007). Neurons in monkey visual area V2 encode combinations of orientations. *Nature Neuroscience*, *10*(10), 1313–1321.

Baker, J. F., Petersen, S. E., Newsome, W. T., & Allman, J. M. (1981). Visual response properties of neurons in four extrastriate visual areas of the owl monkey (Aotus trivirgatus): a quantitative comparison of medial, dorsomedial, dorsolateral, and middle temporal areas. *Journal of Neurophysiology*, *45*(3), 397–416.

Barlow, H. B. (1961). Possible Principles Underlying the Transformations of Sensory Messages.

Bayerl, P., & Neumann, H. (2004). Disambiguating visual motion through contextual feedback modulation. *Neural Computation*, *16*(10), 2041–2066.

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*. http://doi.org/10.1109/72.279181

Blasdel, G., & Fitzpatrick, D. (1984). Physiological organization of layer 4 in macaque striate cortex. *The Journal of Neuroscience*, *4*(3), 880–895.

Blasdel, G., & Salama, G. (1986). © 1986 Nature Publishing Group. *Nature*, *319*(30), 402–403.

Bonhoeffer, T., & Grinvald, a. (1991). Iso-orientation domains in cat visual cortex are arranged in pinwheel-like patterns. *Nature*, *353*(6343), 429–431. http://doi.org/10.1038/353429a0

Bowman, S. R., Manning, C. D., & Potts, C. (2015). Tree-structured composition in neural networks without tree-structured architectures. *arXiv*, 1506.04834.

Callaway, E. M. (1998). Local circuits in primary visual cortex of the macaque monkey. *Annual Review of Neuroscience*, *21*, 47–74.

Callaway, E. M., & Wiser, A. K. (1996). Contributions of individual layer 2-5 spiny neurons to local circuits in macaque primary visual cortex. *Visual Neuroscience*, *13*, 907–922.

Chey, J., Grossberg, S., & Mingolla, E. (1998). Neural dynamics of motion processing and speed discrimination. *Vision Research*, *38*(18), 2769–2786.

Chichilnisky, E. J. (2001). A simple white noise analysis of neuronal light. *Network: Computation in Neural Systems*, *12*, 199–213.

Chubb, C., & Sperling, G. (1988). Drift-balanced random stimuli: a general basis for studying non-Fourier motion perception. *Journal of the Optical Society of America. A, Optics and Image Science*, *5*(11), 1986–2007.

Clark, D. A., Bursztyn, L., Horowitz, M. A., Schnitzer, M. J., & Clandinin, T. R. (2011). Defining the Computational Structure of the Motion Detector in Drosophila.

Neuron, *70*(6), 1165–1177.

Clifford, C. W., Ibbotson, M. R., & Langley, K. (1997). An adaptive Reichardt detector model of motion adaptation in insects and mammals. *Visual Neuroscience*, *14*(4), 741–749.

Clifford, C. W., & Langley, K. (2000). Recursive implementations of temporal filters for image motion computation. *Biological Cybernetics*, *82*(5), 383–390.

Colby, J. G. M. & C. L. (1981). Response properties of single cells in monkey striate cortex during reversible inactivation of individual lateral geniculate laminae. *Journal of Neurophysiology*, *46*(5).

Conway, B. R. (2003). Space-Time Maps and Two-Bar Interactions of Different Classes of Direction-Selective Cells in Macaque V-1. *Journal of Neurophysiology*, *89*(5), 2726–2742.

Cover, T., & Thomas, J. (2012). *Elements of information theory*. New York: John Wiley & Sons, Inc.

Das, A., & Gilbert, C. D. (1999). Topography of contextual modulations mediated by short-range interactions in primary visual cortex. *Nature*, *399*, 655–661.

De Valois, R. L., & Cottaris, N. P. (1998). Inputs to directionally selective simple cells in macaque striate cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(24), 14488–93.

De Weerd, P., Desimone, R., & Ungerleider, L. G. (2003). Impairments in spatial generalization of visual skills after V4 and TEO lesions in macaques (Macaca mulatta). *Behavioral Neuroscience*, *117*(6), 1441–1447.

Dean, A., & Tolhurst, D. (1983). On the distinctness of simple and complex cells in the visual cortex of the cat. *The Journal of Physiology*, (344), 305–325.

Doi, E., & Lewicki, M. S. (2014). A Simple Model of Optimal Population Coding for Sensory Systems. *PLOS Computational Biology*, *10*(8).

Douglas, R. J., Koch, C., Mahowald, M., Martin, K. a, & Suarez, H. H. (1995). Recurrent excitation in neocortical circuits. *Science (New York, N.Y.)*, *269*(5226), 981–5.

Dubner, R., & Zeki, S. M. (1971). Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey. *Brain Research*, *35*(2), 528–532. http://doi.org/10.1016/0006-8993(71)90494-X

Elman, J. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211.

Felleman, D. J., & Kaas, J. H. (1984). Receptive-field properties of neurons in middle temporal visual area (MT) of owl monkeys. *Journal of Neurophysiology*, *52*(3), 488–513.

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

Fitzpatrick, D., Usrey, W. M., Schofield, B. R., & Einstein, G. (1994). The sublaminar organization of corticogeniculate neurons in layer 6 of macaque striate cortex. *Visual Neuroscience*, *11*, 307–315.

Freeman, J., Ziemba, C. M., Heeger, D. J., Simoncelli, E. P., & Movshon, J. A. (2013). A functional and perceptual signature of the second visual area in primates. *Nature Neuroscience*, *16*(7), 974–81.

Funahashi, K., & Nakamura, Y. (1993). Approximation of dynamical systems by continuous time recurrent neural networks. *Neural Networks*, *6*(6), 801–806.

Hateren, J. H. Van. (1992). A theory of maximizing sensory information. *Biological Cybernetics*, *68*, 23–29.

Hegdé, J., & Van Essen, D. C. (2000). Selectivity for complex shapes in primate visual area V2. *The Journal of Neuroscience*, *20*(5), RC61.

Hegdé, J., & Van Essen, D. C. (2004). Temporal dynamics of shape analysis in macaque visual area V2. *Journal of Neurophysiology*, *92*(5), 3030–3042.

Hermundstad, A. M., Briguglio, J. J., Conte, M. M., & Victor, J. D. (2014). Variance predicts salience in central sensory processing. *eLife*, *3*, 1–40.

Heydt, R. von der, & Peterhans, E. (1989). Mechanisms of Contour Perception in Monkey Visual Cortex. I. Lines of Pattern Discontinuity. *The Journal of Neuroscience*, *9*(5), 1731–1748.

Heydt, R. Von Der, Peterhans, E., & Baumgartner, G. (1984). Illusory Contours and Cortical Neuron Responses. *Science*, *224*(4654), 1260–1262.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1–32.

Hornik, K., Stinchcombe, M., & White, H. (1988). Multilayer feedforward networks are universal approximators. *Neural Networks, 2*, 359–366.

Hubel, D. H., & Livingstone, M. S. (1987). Segregation of Form, Color and Stereopsis in Primate Area 18. *The Journal of Neuroscience*, *7*(11), 3378–3415.

Hubel, D. H., & Wiesel, T. N. (1974). Sequence regularity and geometry of orientation columns in the monkey striate cortex. *The Journal of Comparative Neurology*, *158*(3), 267–293.

Hubel, D., & Wiesel, T. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, (160), 106–154.

Hubener, M., & Bolz, J. (1992). Relationships between Dendritic Morphology and Cytochrome-Oxidase Compartments in Monkey Striate Cortex. *Journal of Comparative Neurology*, *324*(1), 67–80.

Ito, M., & Komatsu, H. (2004). Representation of Angles Embedded within Contour Stimuli in Area V2 of Macaque Monkeys. *The Journal of Comparative Neurology*, *24*(13), 3313–3324.

Joukes, J., Hartmann, T. S., & Krekelberg, B. (2014). Motion detection based on recurrent network dynamics. *Frontiers in Systems Neuroscience, 8*(December), 239.

Julesz, B. (1981). Textons, the elements of texture perception, and their interactions. *Nature, 290*.

Kaas, J. O. N. H. (1971). A representation of the visual field in the caudal third of the middle temporal gyrus of the owl monkey. *Brain Research, 31*, 85–105.

Kalchbrenner, N., Danihelka, I., & Graves, A. (2015). Grid Long Short-Term Memory. *arXiv*, 1–14.

Katz, C., Gilbert, C. D., & Wiesel, T. N. (1989). Local Circuits Cortex. *The Journal of Neuroscience*, *9*(April), 1389–1399.

Koenderink, J. J., van Doorn, a J., & van de Grind, W. a. (1985). Spatial and temporal parameters of motion detection in the peripheral visual field. *Journal of the Optical Society of America. A, Optics and Image Science*, *2*(2), 252–9.

Kohn, A., & Movshon, J. A. (2003). Neuronal adaptation to visual motion in area MT of the macaque. *Neuron*, *39*(4), 681–91.

Kourtzi, Z., & Kanwisher, N. (2000). Implied motion activates extrastriate motion-processing areas: Response to David and Senior (2000). *Trends in Cognitive Sciences*, *4*(8), 295–296.

Kourtzi, Z., Krekelberg, B., & van Wezel, R. J. A. (2008). Linking form and motion in the primate brain. *Trends in Cognitive Sciences*, *12*(6), 230–236.

Krekelberg, B. (2008). Motion detection mechanisms. *The Senses: A Comprehensive Reference*, 133–155.

Krekelberg, B., Dannenberg, S., Hoffmann, K.-P., Bremmer, F., & Ross, J. (2003). Neural correlates of implied motion. *Nature*, *424*, 674–677.

Krekelberg, B., van Wezel, R. J. a, & Albright, T. D. (2006). Adaptation in macaque MT reduces perceived speed and improves speed discrimination. *Journal of Neurophysiology*, *95*(1), 255–70. http://doi.org/10.1152/jn.00750.2005

Krekelberg, B., Vatakis, A., & Kourtzi, Z. (2005). Implied Motion From Form in the Human Visual Cortex. *Journal of Neurophysiology*, *94*(6), 4373–4386.

Lang, K. J., Waibel, A. H., & Hinton, G. E. (1990). A time-delay neural network architecture for isolated word recognition. *Neural Networks*, *3*(1), 23–43.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2323.

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, *20*(7), 1434–48. http://doi.org/10.1364/JOSAA.20.001434

Lee, T. S., & Nguyen, M. (2001). Dynamics of subjective contour formation in the early visual cortex. *PNAS*, *98*(4).

Liang, M., & Hu, X. (2015). Recurrent Convolutional Neural Network for Object Recognition. *Cvpr*, 3367–3375.

Liao, Q., & Poggio, T. (2016). Bridging the Gaps Between Residual Learning, Recurrent Neural Networks and Visual Cortex. *arXiv Preprint*, (047), 1–16.

Livingstone, M. S., Pack, C. C., & Born, R. T. (2001). Two-dimensional substructure of MT receptive fields. *Neuron*, *30*(3), 781–793.

Livingstone, S., & Hubel, H. (1984). Anatomoy and physiology of a color system in the primate visual cortex. *The Journal of Neuroscience*, *4*(1), 309–356.

Maex, R., & Orban, G. (1996). Model circuit of spiking neurons generating directional selectivity in simple cells. *Journal of Neurophysiology*, *75*(4).

Malach, R. (1992). Dendritic sampling across processing streams in monkey striate cortex. *J Comp Neurol*, *315*(3), 303–312.

Malach, R., Amir, Y., Harel, M., & Grinvald, A. (1993). Relationship between intrinsic connections and functional architecture revealed by optical imaging and in vivo targeted biocytin injections in primate striate cortex. *Proc. Natl. Acad. Sci. U S A*, *90*(22), 10469–10473.

Marmarelis, P. Z., & Marmarelis, V. Z. (1978). *Analysis of physiological systems*. Boston, MA: Springer US.

Maunsell, J. H. F., & Nealey, T. A. (1990). Magnocellular and Parvocellular Contributions to Responses in the Middle Temporal Visual Area (MT) of the Macaque Monkey. *The Journal of Neuroscience*, (October), 3323–3334.

Maunsell, J. H., & van Essen, D. C. (1983). The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *3*(12), 2563–2586.

McKee, S. P., & Welch, L. (1985). Sequential recruitment in the discrimination of velocity. *Journal of the Optical Society of America. A, Optics and Image Science*, *2*(2), 243–51.

Mechler, F., & Ringach, D. L. (2002). On the classification of simple and complex cells. *Vision Research*, *42*(8), 1017–33.

Merigan, W. H., & Maunsell, J. H. R. (1993). How parallel are the primate visual pathways? *Annu. Rev. Neurosci.*, *16*, 369–402.

Mikami, a. (1992). Spatiotemporal characteristics of direction-selective neurons in the middle temporal visual area of the macaque monkeys. *Experimental Brain Research. Experimentelle Hirnforschung. Expérimentation Cérébrale*, *90*(1), 40–6.

Mineiro, P., & Zipser, D. (1998). Analysis of direction selectivity arising from recurrent cortical interactions. *Neural Computation*, *10*(2), 353–371.

Movshon, J. a, & Newsome, W. T. (1996). Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. *The Journal of Neuroscience*, *16*(23), 7733–41.

Movshon, J. A., Thompson, I. D., & Tolhurst, D. J. (1978). Receptive field organization of complex cells in the cat's striate cortex. *Journal of Physiology*, *283*, 79–99.

Nealey, A., & Maunsell, H. R. (1994). Magnocellular and parvocellular contributions to the responses of neurons in macaque striate cortex. *The Journal of Neuroscience*, *14*(April), 2069–2079.

Nenadic, Z., & Burdick, J. W. (2005). Spike detection using the continuous wavelet transform. *Ieee Tbme*, *52*(1), 74–87.

Orban, G. a, Fize, D., Peuskens, H., Denys, K., Nelissen, K., Sunaert, S., … Vanduffel, W. (2003). Similarities and differences in motion processing between the human and macaque brain: Evidence from fMRI. *Neuropsychologia*, *41*(13), 1757–1768.

Orban, G. a. (1986). Velocity sensitivity and direction selectivity of neurons in areas V1 and V2 of the monkey: influence of eccentricity. *Journal of Neurophysiology*, *56*(2).

Pillow, J. W., & Simoncelli, E. P. (2006). Dimensionality reduction in neural models: an information-theoretic generalization of spike-triggered average and covariance analysis. *Journal of Vision*, *6*(4), 414–28.

Priebe, N. J., & Ferster, D. (2005). Direction selectivity of excitation and inhibition in simple cells of the cat primary visual cortex. *Neuron*, *45*(1), 133–45.

Qiu, F. T., & Heydt, R. Von Der. (2005). Figure and Ground in the Visual Cortex : V2 Combines Stereoscopic Cues with Gestalt Rules. *Neuron*, *47*, 155–166.

Rust, N. C., Schwartz, O., Movshon, J. A., & Simoncelli, E. (2004). Spike-triggered characterization of excitatory and suppressive stimulus dimensions in monkey V1. *Neurocomputing*, *58-60*, 793–799.

Rust, N. C., Schwartz, O., Movshon, J. A., & Simoncelli, E. P. (2005). Spatiotemporal elements of macaque v1 receptive fields. *Neuron*, *46*(6), 945–56.

Sabatini, S. P., & Solari, F. (1999). An architectural hypothesis for direction selectivity in the visual cortex: the role of spatially asymmetric intracortical inhibition. *Biological Cybernetics*, *80*(3), 171–83.

Salinas, E., & Abbott, L. F. (1996). A model of multiplicative neural responses in parietal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(21), 11956–11961.

Sawatari, A., & Callaway, E. M. (2000). Diversity and Cell Type Specificity of Local Excitatory Connections to Neurons in Layer 3B of Monkey Primary Visual Cortex.

Neuron, *25*, 459–471.

Schlack, A., Krekelberg, B., & Albright, T. D. (2007). Recent history of stimulus speeds affects the speed tuning of neurons in area MT. *The Journal of Neuroscience*, *27*(41), 11009–18.

Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, *45*(11), 2673–2681.

Simoncelli, E., & Heeger, D. (1998). A model of neuronal responses in visual area MT. *Vision Research*, *38*(5), 743–761.

Simoncelli, E. P., Paninski, L., Pillow, J., & Schwartz, O. (2004). *Characterization of Neural Responses with Stochastic Stimuli*. *The new cognitive Neurosciences*. MIT press.

Sincich, L. C., & Horton, J. C. (2002). Divided by Cytochrome Oxidase : A Map of the Projections from V1 to V2 in Macaques. *Science*, *295*(March), 1734–1738.

Sincich, L. C., & Horton, J. C. (2005). THE CIRCUITRY OF V1 AND V2: Integration of Color, Form, and Motion. *Annual Review of Neuroscience*, *28*(1), 303–326.

Sincich, L. C., Park, K. F., Wohlgemuth, M. J., & Horton, J. C. (2004). Bypassing V1: a direct geniculate input to area MT. *Nature Neuroscience*, *7*(10), 1123–1128.

Skottun, B., Valois, R. De, & Grosof, D. (1991). Classifying simple and complex cells on the basis of response modulation. *Vision Research*, *31*(7), 1079–1086.

Suarez, H., Koch, C., & Douglas, R. (1995). Modeling direction selectivity of simple cells in striate visual cortex within the framework of the canonical microcircuit. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *15*(10), 6700–19.

Sutton, R. S., Werbos, P. J., Gupta, N. K., Rosenfeld, E., Mccool, J., & Jolla, L. (1988). Beyond regression: new tools for prediction and analysis in the behavioral sciences.

Tkacik, G., Prentice, J. S., Victor, J. D., & Balasubramanian, V. (2010). Local statistics in natural scenes predict the saliency of synthetic textures. *PNAS*, *107*(42), 18149–18154.

Tlapale, E., Masson, G. S., & Kornprobst, P. (2010). Modelling the dynamics of motion integration with a new luminance-gated diffusion mechanism. *Vision Research*, *50*(17), 1676–1692.

Ts’o, D. Y., Gilbert, C. D., & Wiesel, T. N. (1986). Relationships between horizontal interactions and functional architecture in cat striate cortex as revealed by cross-correlation analysis. *The Journal of Neuroscience*, *6*(4), 1160–1170.

Ungerleider, L. G., & Desimone, R. (1986). Cortical connections of visual area MT in the macaque. *The Journal of Comparative Neurology*, *248*(2), 190–222.

Valois, R. L. De, Cottaris, N. P., Mahon, L. E., Elfar, S. D., & Wilson, J. A. (2000). Spatial and temporal receptive fields of geniculate and cortical cells and directional selectivity. *Vision Research*, *40*, 3685–3702.

Victor, J. D., & Conte, M. M. (1991). Spatial organization of nonlinear interactions in form perception. *Vision Research*, *31*(9), 1457–1488.

Victor, J. D., & Conte, M. M. (2012). Local image statistics: maximum-entropy constructions and perceptual salience. *J Opt Soc Am A Opt Image Sci Vis*, *29*(7), 1313–1345.

Warner, C. E., Goldshmit, Y., & Bourne, J. a. (2010). Retinal afferents synapse with relay cells targeting the middle temporal area in the pulvinar and lateral geniculate nuclei. *Frontiers in Neuroanatomy*, *4*(February), 8.

Watson, A., & Ahumada, A. (1985). Model of human visual-motion sensing. *Optical Society of America*, *2*(2).

Yu, Y., Schmid, A. M., & Victor, J. D. (2015). Visual processing of informative multipoint correlations arises primarily in V2. *eLife*, *4*, 1–13.

Zeki, S. M. (1974). Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. *The Journal of Physiology*, *236*(3), 549–73.