Simulating component-to-pattern dynamic effects with a computer model of middle temporal pattern neurons

John A. Perrone

The School of Psychology, University of Waikato, Hamilton, New Zealand

Richard J. Krauzlis

Laboratory of Sensorimotor Research, NEI, NIH, Bethesda, MD, USA

Some primate motion-sensitive middle temporal (MT) neurons respond best to motion orthogonal to a contour's orientation (component types) whereas another class (pattern type) responds maximally to the overall pattern motion. We have previously developed a model of the pattern-type neurons using integration of the activity generated in speed- and direction-tuned subunits. However, a number of other models have also been able to replicate MT neuron pattern-like behavior using a diverse range of mechanisms. This basic property does not really challenge or help discriminate between the different model types. There exist two sets of findings that we believe provide a better yardstick against which to assess MT pattern models. Some MT neurons have been shown to change from component to pattern behavior over brief time intervals. MT neurons have also been observed to switch from component- to pattern-like behavior when the intensity of the intersections in a plaid pattern stimulus changes. These properties suggest more complex time- and contrast-sensitive internal mechanisms underlying pattern motion extraction, which provide a real challenge for modelers. We have now replicated these two component-to-pattern effects using our MT pattern model. It incorporates two types of V1 neurons (sustained and transient), and these have slightly different time delays; this initially favors the component response, thus mimicking the temporal effects. We also discovered that some plaid stimuli contain a contrast asymmetry that depends on the plaid direction and the intensity of the intersections. This causes the model MT pattern units to act as component units.

Introduction

The middle temporal (MT/V5) area is one of the most studied visual areas of the brain. Given its key role in visual motion processing and the predominance of visual motion in our environment, it is not surprising that MT neurons have been extensively tested and a wide range of response properties uncovered (for reviews see Born & Bradley, 2005; Bradley & Goyal, 2008; Krekelberg & Albright, 2005; Newsome, Britten, Salzman, & Movshon, 1990; Pack & Born, 2008). Numerous attempts have also been made to build theoretical models of the mechanisms underlying MT neuron properties (e.g., Adelson & Movshon, 1982; Albright, 1984; Bowns, 2002; Chey, Grossberg, & Mingolla, 1997; Grzywacz & Yuille, 1990; Johnston, McOwan, & Buxton, 1992; Movshon, Adelson, Gizzi, & Newsome, 1985; Nishimoto & Gallant, 2011; Nowlan & Sejnowski, 1995; Perrone, 2004; Qian, Andersen, & Adelson, 1994; Rust, Mante, Simoncelli, & Movshon, 2006; Simoncelli & Heeger, 1998; Snowden, Treue, Erickson, & Andersen, 1991), but models that can use two-dimensional image sequences for input and which have realistic front-end filters that match the properties of the neurons in the stage preceding MT (V1) are few and far between (see Perrone, 2004). In order to correctly compare such models against MT neurons, the models must be able to be tested with the same stimuli used to test the neurons so that factors such as contrast and spatial frequency content are taken into account.

Over the years, we have developed a model of MT neurons that can be tested with real image sequences and that is able to simulate the basic properties of these neurons. The response to multiple directions of motion (Adelson & Movshon, 1982; Albright, 1984) and speed (Felleman & Kaas, 1984; Lagae, Raiguel, & Orban, 1993; Maunsell & Van Essen, 1983; Rodman & Albright, 1987) has been replicated using this model (Perrone, 2004, 2005; Perrone & Thiele, 2002). MT neuron responses to different spatial and temporal


doi: 10.1167/14.1.19

Received July 1, 2013; published January 22, 2014

ISSN 1534-7362 © 2014 ARVO
frequencies (Perrone & Thiele, 2001; Priebe, Cassanello, & Lisberger, 2003) have also been simulated (Perrone, 2005; Perrone & Thiele, 2002). Spatial effects that result from the location of small patches of moving gratings ("pseudoplaids") on the direction tuning of MT neurons (Majaj, Carandini, & Movshon, 2007) have also been replicated using our MT pattern neuron model (Perrone & Krauzlis, 2008). Similarly, the spatial inhibitory surround effects that have been discovered outside the classical receptive fields of MT neurons (Xiao, Raiguel, Marcar, & Orban, 1997) can be explained (Perrone, 2012) using inhibitory inputs to our model MT pattern units designed to reduce the amount of redundant velocity signals passed onto the next stage of visual motion processing (medial superior temporal).

All in all, we have been able to replicate many MT neuron properties using computer software that starts with a relatively simple front-end spatiotemporal filtering stage (based on V1 neuron properties) and which combines the signals from these V1-like filters in a straightforward way (mainly simple integration) (see Figure 1). Many different designs have been proposed for how MT pattern neurons derive their ability to respond mainly to the overall direction of a moving pattern (e.g., Adelson & Movshon, 1982; Alais, Wenderoth, & Burke, 1994; Albright, 1984; Bowns, 2002; Chey et al., 1997; Okamoto et al., 1999; Perrone & Krauzlis, 2008; Rust et al., 2006; Sereno, 1993; Wilson, Ferrera, & Yo, 1992). Despite containing quite different internal mechanisms underlying how the pattern behavior is generated, these models can all demonstrate pattern responses that generally agree with those found in actual MT neurons. The basic test of direction tuning does not challenge these models sufficiently to reveal any deficiencies in their design or to highlight any mismatches between the model components and the actual neural mechanisms.

There are two prominent studies of MT neurons that provide a better challenge for motion models and that seem to require greater sophistication in the model design in order to replicate the effect in computer software. Pack and Born (2001) showed that MT neurons initially respond primarily to the component of motion perpendicular to a contour’s orientation, but over a brief period (approximately 100 ms), the responses gradually change to indicate the true stimulus direction, regardless of its orientation. This property suggests the involvement of recurrent or feedback processes to compute motion direction, and this is what was proposed by Pack and Born (2001) along with the idea that the visual system computes the motion of contours and endpoints via different pathways.

A second study that provides a stronger challenge to motion models was carried out by Stoner and Albright (1992). They found that some MT neurons change their direction tuning properties when the regions of overlap of two superimposed moving gratings is altered to correspond to transparent gratings overlaying one another. This MT property replicated perceptual effects previously demonstrated by Stoner, Albright, and Ramachandran (1990) with human observers. The observers perceived either a single coherently moving "plaid pattern" or two component gratings sliding noncoherently across one another, depending on the luminance of the regions of overlap. The amount of motion integration that occurred depended somehow on the "transparency" that was present in the stimulus, and Stoner and Albright (1992, 1993) argued that this "selective integration" was achieved by the visual system using segmentation cues to classify motion signals according to physical origin.

Both of these data sets provide a strong challenge to computer models of MT pattern neurons and require a mechanism that is time-sensitive and which can be influenced by the luminance level of the plaid intersections. We discovered that the design of our model pattern neurons has the required properties to accommodate both of these phenomena. Our model MT neurons (Figure 1) consist of subunits constructed from spatiotemporal filters based on V1 neurons (Perrone &
The front end of the model takes movie sequences that are convolved with two types of spatiotemporal energy filters (sustained and transient, S and T) based on V1 neurons with low-pass and band-pass temporal frequency tuning (Foster, Gaska, Nagler, & Pollen, 1985; Hawken, Shapley, & Grosof, 1996). The S and T energy is combined using the weighted intersection mechanism (WIM) proposed by Perrone and Thiele (2002). This stage introduces tight speed tuning to a set of intermediate sensors that form the subunits of our MT model neurons. These WIM subunits are considered to be analogs of the speed-tuned V1 neurons discovered by Priebe, Lisberger, and Movshon (2006) and were designed to replicate the spatiotemporal frequency-tuning properties of these cells (see spatial frequency-temporal frequency contour plot in Figure 1).

The WIM subunits sample multiple directions and speeds across a small area of the input image sequence (Perrone, 2004; Perrone & Krauzlis, 2008) that represents the receptive field of our MT neuron (Figure 1).

Some subunits are tuned to directions away from the primary tuning of the MT unit (referred to as “off-axis” units. See small arrows in the WIM cluster depicted in Figure 1) and are tuned to a slower speed than the units tuned to the primary direction (Perrone, 2005). The speed is a function of the cosine of the angle between the direction tuning of the WIM subunit and the primary direction of the MT unit. These are designed to pick up the motion from other edge orientations making up the moving pattern (Figure 2). A moving 2-D shape on the retina containing multiple edge orientations will maximally stimulate one of our MT pattern units that is tuned to the overall direction and speed of the object. The different edge orientations making up the object shape will have a speed that is a (cosine) function of the angle between the edge orientation and the object’s direction of motion. The speed tuning of the MT pattern model subunits is set to match these different possible edge speeds. For the MT unit to be sufficiently selective for a particular direction and speed, the subunits need to be tightly tuned for speed (Perrone, 2004). Figure 2c shows the tight speed tuning that results from the WIM model stage. The broadly tuned temporal frequency tuning of the S and T V1 inputs (Figure 2b) is converted to tight speed tuning via the WIM stage operator. The WIM stage can be conceptualized as a type of “AND” operator that produces a large output whenever both the S and T signals are high and equal (Equation 1). The peak response occurs when the two temporal frequency curves cross and the resulting WIM unit speed tuning curve is tightly tuned (Figure 2c). The WIM stage also provides an economical means (Perrone, 2005) of creating the cosine-weighted speed tuning values required to construct the pattern units (Figure 2d, e). The inclusion of subunits with tight speed tuning makes our model MT units unique. They do not calculate the intersection of constraints (Adelson & Movshon, 1982; Okamoto et al., 1999; Sereno, 1993) and do not assume that velocity information is available at the level of MT as is the case for techniques based on vector averaging or feature tracking (Alais et al., 1994; Bowns, 2002; Wilson et al., 1992). This arrangement of WIM units enables the overall MT pattern motion detector (PMD) to respond maximally to a moving texture or shape containing multiple edge orientations moving in its preferred direction.

The total activity of the model PMD is based on simple integration of the responses across the set of WIM subunits. The model uses opponency from WIM units tuned to directions opposite to the preferred direction of the unit (red arrows in Figures 1 and 2a), but this is subtracted from each of the WIM clusters (flowerets in Figure 2a) prior to the integration (Perrone & Krauzlis, 2008). The model does not incorporate any feedback from higher levels or use “nonmotion” input signals from visual areas outside of V1 (e.g., V2). On the surface, therefore, it seems unlikely that such a model would be able to reproduce the effects discovered by Pack and Born (2001) and Stoner and Albright (1992). It lacks the features that these authors proposed as the underlying mechanisms behind their results.

We have since discovered that these two MT phenomena are not beyond the scope of our basic MT neuron model. We have been able to replicate both the Pack and Born (2001) and the Stoner and Albright (1992) results using our basic MT neuron pattern motion model (Perrone, 2004; Perrone & Krauzlis, 2008). The temporal effects discovered by Pack and Born (2001) can be explained by the fact that the output of our MT model units is controlled by the relative activity of the S and T V1-stage spatiotemporal filters that make up the WIM subunits (Figure 1b through e and Equation 1, Methods). We have also been able to replicate another temporal component-to-pattern effect discovered by Smith, Majaj, and Movshon (2005) by introducing a time delay in the arrival of the inhibition from the opponent WIM units. The Stoner and Albright (1992) effect can be traced to local asymmetries in the contrast of the plaids and to the existence of the off-axis WIM subunits in our model. We show that a similar mechanism can also explain the response properties of MT neurons when tested with “depth-ordered plaids” (Thiele & Stoner, 2003).

Here we describe the simulation procedure and the underlying mechanisms in the model that produce the particular MT neuron properties. We conclude that the Pack and Born (2001), Smith et al. (2005), Stoner and Albright (1992), and Thiele and Stoner (2003) effects...
can all be explained by our MT pattern neuron model, and we have raised the bar as to the range of MT neuron data that must now be replicated by models claiming to be analogs of MT pattern neurons.

**Methods**

**Model MT units**

Our model PMD units are designed to mimic MT pattern neurons (Albright, 1984; Movshon et al., 1985) and consist of subunits based on V1 spatiotemporal filters (Figure 1). The details of the stages leading up to the MT model units have been outlined in detail previously (Perrone, 2004, 2005; Perrone & Krauzlis, 2008; Perrone & Thiele, 2002). The spatiotemporal filters at the initial filtering stage of the model are based on the temporal (Foster et al., 1985; Hawken et al., 1996) and spatial (Hawken & Parker, 1987) frequency tuning of V1 neurons.

The spatiotemporal energy outputs of the S and T filters are combined using the WIM model proposed by Perrone and Thiele (2002), which was designed to produce the maximum output from a combination of the two S & T filter inputs whenever their output is both high and equal. This creates sensors with tight speed tuning (Figure 2c). Here we follow Perrone (2012) and use a slightly modified version of the original (Perrone & Thiele, 2002) WIM model equation:

\[
\text{WIM} = \frac{S' + T'}{|S' - T'| + \delta}
\]  

(1)

where \(\delta\) is the delta term used in the original equation and which controls the bandwidth of the speed tuning of the WIM sensor. It was set to a value of 8.0 in all of the simulations reported in this paper. The \(S'\) and \(T'\) indicate that the original spatiotemporal energy from the S and T filters is transformed using a gain control mechanism (Perrone, 2012).

\[
S' = \frac{aS}{pS + sc}
\]  

(2)

\[
T' = \frac{aT}{pT + tc}
\]  

(3)

where \(a = 6.8, p = .06, sc = .15\) and \(tc = 0.14\) (for S and T values in the range 0–60).

In the remainder of the text, we will continue to refer to this transformed energy as S and T to simplify the nomenclature. Equation 1 shows that the WIM output is based on the total amount of S and T energy (numerator of Equation 1), and it also indicates that, when the S and T values are similar, the WIM output is high, and when they are dissimilar, the output is low (denominator of Equation 1). The relative amounts of energy in the S and T spatiotemporal filters determines
the output of the WIM filters, and this, in turn, controls the level of the MT PMD unit.

Each cluster within a PMD is made up from seven “positive” WIM subunits and five inhibitory (opponent) subunits (black and red arrows in Figure 2a). Their direction tuning ranges from 0° to 330° in 30° steps. The speed tuning is a cosine function of the difference between the direction tuning value and the optimum overall direction tuning (θ) for the PMD. Therefore, if the overall velocity tuning of the PMD is \( V_p = (V_p, \theta_p) \), then the speed tuning of the cluster subunits making up the detector in the model is given by \( s_i = V_p \cos(\theta_i - \beta_i) \) where \( \beta_i \) ranges from 0° to 330° in 30° steps. The set of \( \beta_i \) values was designed to sample the range of possible edge orientations that could be present in the receptive field of the PMD, and it sets up the speed tuning of each cluster subunit to match the expected speed of the different possible edge configurations. The slower speed values for the off-axis units are obtained by a simple weighting of the T output relative to the S output (Perrone, 2005). For example, the −60° unit in Figure 2a is tuned to half the speed of the 0° unit by an increase in the gain of the T unit (Figure 2d, e).

The WIM sensor clusters are spatially separated in a circular array (Figures 1 and 2a). The radial separation distance between clusters depends on the spatial frequency tuning (\( u_0 \)) of the WIM subunits (2 c/° for our 2°/s MT units) and was set at a distance \( d = \frac{8}{u_0} \) pixels for the simulations reported in this paper. The differently (\( s_i, \beta_i \)) tuned WIM subunits are weighted prior to their output being summed across all of the clusters. For \( \beta \) values ±30° on either side of \( \theta_p \), \( w_i = 0.87 \); for values ±60°, \( w_i = 0.5 \), and for those ±90°, \( w_i = 0.3 \). Subunits that are tuned to directions within the range \( \theta_p - 180° \pm 60° \) (red lines in Figure 2a) contribute in an “opponent” fashion (\( w = -1.0 \)). The output from the opponent units in a cluster is subtracted from the positive activity in the same cluster. The net local output (cluster positive activity minus cluster negative activity) from all of the nine clusters in the receptive field is half-wave rectified, then summed (Perrone & Krauzlis, 2008). This is the output we report on in the simulations, and it is intended to correspond to the firing rates of MT neurons.

To test the model units, we used image sequences (128 × 128 pixels × 8 frames) that correspond to a 266-ms time sample and 4.25° × 4.25° of visual angle. We report measurements in pixels because this is the natural unit for digital images, but we report speeds in degree per second (1 pixel/frame corresponds to 1°/s). The MT model unit we tested for both the Pack and Born (2001) and the Stoner and Albright (1992) simulations was tuned to 2°/s and 0° direction. All parameters in the model were kept the same for the two tests.

### Stimuli

For the Pack and Born (2001) tests, we used input movie sequences consisting of five short line segments (8 pixels long and 3 pixels wide) in an array with a central line and four flanking lines with their centers offset ±16 pixels relative to the center unit (See Figure 3a). Following Pack and Born (2001), the small bar array was oriented at either 0°, −45°, or +45° relative to the direction of motion. The array moved in eight different directions (0° to 315° in 45° steps) at a speed of 2°/s.

For the Stoner and Albright (1992) simulations, we matched their stimuli as closely as possible and used plaid patterns (see Figure 6a) made up of two square wave gratings (135° separation angle, .28 duty cycle, and three cycles per image width = .71 c/°). The pixel intensity of the background was 220; the grating bars were 116; and the intersections were 12, 76, or 136. These values are proportional to the grating luminance values used by Stoner and Albright (1992)—55, 29 for the background and bars, respectively, and 3, 19, and 34 cd/m² for the intersections—and produce the same luminance ratios. We will use the label “transparent” to signify the intermediate intersection intensity region that tends to produce noncoherent plaid responses in human observers (Stoner et al., 1990) and which produced component-like behavior in MT neurons (Stoner & Albright, 1992). The MT model unit was tested with three directions of plaid motion (−67.5°, 0°, and 67.5°) at the three different intersection intensity values, giving a set of nine different output values. The MT model unit was located at image location (48, 63). This was established by finding the MT unit across the image that produced the best match to the Stoner and Albright neuron. Following Stoner and Albright (1992), the model pattern units were also tested with simple sine gratings (2 c/°, 100% contrast) moving at a range of directions from −135° to 135° in 67.5° steps.

### Results

#### Temporal effects

**Pack and Born (2001) simulations**

Figure 3 shows the results of testing one of our model pattern neurons (see Methods) tuned to 0° direction and 2°/s with an array of small line segments similar to the stimuli used by Pack and Born (2001). The model uses an eight-frame sequence to extract a signal, and we recorded the output of the different stages across each of these frames. Because of wraparound effects caused by convolution of the spatiotemporal filters, with the input image sequences, the first and last frame output tends to
be noisy and low in magnitude. We therefore report mainly on frames 2–6 in our analysis of the temporal evolution of the MT unit signal. The “early” response reported by Pack and Born (2001) was taken as the model unit output at frame 2 (66.5 ms). The “time averaged” output was the average across frames 3–6 (100–200 ms). The different orientations of the line segments used in the tests are shown in Figure 3a for the case in which the pattern motion was to the right (0°). As in Pack and Born (2001), the angle of the bars relative to the direction of movement (Φ) was varied: In one condition, the line segments were orthogonal to the motion direction (Φ = 90°); in another, it was tilted −45° to the direction of motion (Φ = 45°), and in the other +45° (Φ = 135°). Figure 3b shows in polar plot form the output of the model MT unit in response to a range of directions at frame 2 of the response profile. Pack and Born (2001) calculated the preferred direction (PD) for each of their cells by finding the vector average of the stimulus direction weighted by the response to that direction. We calculated the same measure for each of our data sets.
When the line segments were oriented orthogonal to the direction of motion (red curve), the peak response occurred in the direction corresponding to the direction tuning of the unit (PD = 0.55°). However, when the bars were tipped −45° relative to the pattern direction (blue curve), the peak response occurred for the 45° direction in the early stage of the MT model unit response. Similarly, for the φ = 135° case (green curve), the peak response was in the 315° direction. The PD was 21.5° for φ = 45° and −23.5° for 135° (dashed lines). These are “component” responses because the unit is not responding best to the overall direction of the pattern of bars. The model MT unit’s response is influenced by the orientation of the bars (see blue and green curves in Figure 3b).

However, when the average response is taken over frames 2–6 (100–200 ms) of the response profile, the MT unit now starts to act as a true pattern motion
detector, and the peak response occurs when the direction of motion aligns with the PD of the unit independent of the small line segment orientation (Figure 3c). This behavior mimics the temporal dynamics of some MT neurons discovered by Pack and Born (2001, see their figure 2).

Pack and Born (2001) calculated the PD relative to the time averaged, \( \phi = 90^\circ \), response and plotted it for different times after the onset of the stimulus motion. Their data from 60 MT neurons is shown in Figure 3d. We carried out a similar analysis on our model pattern unit, and the three curves for \( \phi = 90^\circ \) (red), \( \phi = 45^\circ \) (blue), and \( \phi = 135^\circ \) (green) are shown in Figure 3e. The change from component-type behavior to pattern behavior apparent over time in the Pack and Born (2001) data set is replicated in the behavior of our model MT unit.

In order to understand why the model units displayed this behavior, we analyzed the responses of the input units (the V1 spatiotemporal filters and WIM subunits) that feed into the MT model unit (Methods). We discovered that the sustained spatiotemporal energy units (S) take longer to develop their maximum output compared to the transient units (T). The reason for this can be seen in Figure 4a and b, which shows the space-time plots of the V1-like spatiotemporal filters used in our model. The S type (Figure 4a) are nonoriented in this type of plot and so are nondirectional whereas the T types (Figure 4b) have space-time orientation and are directional (Adelson & Bergen, 1985; Watson & Ahumada, 1985). Both of these filter types occupy the same \((x, y)\) location in the array of filters processing the image. It is possible to represent
the location of a moving edge in this type of plot with an oriented line (white line in Figure 4a, b). Because of its lack of orientation, the excitatory region of the S filters takes longer to be exposed to the moving edge than do the T filters (see dashed arrows).

This is supported by measurements of the energy from each of these two types of filters in response to the line-segment stimuli used by Pack and Born (2001). Figure 4c is for the case in which the bars are oriented $-45^\circ$ to the direction of movement ($0^\circ$). The first graph shows the S and T energy values at different time intervals from the start of the motion sequence. This is for an MT unit located at the middle of the image, and the measured WIM subunit and its spatiotemporal filter inputs are also at this location. As predicted from the spatiotemporal plots (Figure 4a), the S energy takes longer to grow than the T energy and is below the T level for the first 67 ms (frames 1 and 2). Around 100 ms from the start of the motion (frame 3), the two are close to being equal, and so the WIM sensor output peaks at this point. The WIM sensors generate the most output when the S and T inputs are both high and equal (Equation 1, Methods).

The insets above the S and T plots show the WIM activity for a cluster at the center of the MT receptive field. The plots show the amount of WIM sensor activity for each of the 12 angles in the cluster (the maximum output depicted is equal to 22 units of activity). For the Figure 4c case, the most activity is generated in the WIM sensor tuned to $0^\circ$. The WIM responses for frame 2 (67 ms) are shown on the left, and the responses for frame 6 (200 ms) are shown on the right. For the $0^\circ$ direction case, the WIM output is less for frame 2 than it is for frame 6. The model MT unit sums the WIM output from nine different WIM clusters (Figures 1 and 2a), but the dominance of the frame 6 response shown for the center location is also apparent at the other cluster locations (now shown). The MT unit (containing the WIM subunits) therefore has an output that grows from frame 2 (67 ms), peaks around frame 3 (100 ms), and then levels off for the 130–200 ms section of the temporal response profile (rightmost panel in Figure 4c). We would also expect a brief peak in response when the two curves cross again (outside of our measurement window), but this would be short compared to the time that the two curves follow each other in amplitude. The black arrow to the left (marked as circle 1a) indicates the value for the early-stage response, and the gray arrow (1b circle) is the time-averaged response. The latter is larger than the former for the case of $0^\circ$ motion.

For the case in which the line segment pattern moves $45^\circ$ to the PD of the MT model unit (Figure 4d), the evolution of the MT unit output is different. The speed of the motion in the $0^\circ$ direction is now slower than the optimum speed tuning of the MT unit ($2^\circ$/s). It is equal to $2\cos(45^\circ) = 1.4^\circ$/s, which alters the relative sizes of the S and T energy outputs (first graph in Figure 4d). Again the S output builds from frame 1, but this time, because the bars are moving slower than the optimum speed of the WIM and MT units (tuned so that S and T are approximately equal when the speed is $2^\circ$/s), the S output reaches a higher value than T. This means that the WIM subunits peak as the S and T outputs cross (around frame 2 = 67 ms) but then drop their activity as time progresses. This is because the difference between S and T increases as time progresses (see Equation 1, Methods). The S and T values eventually get closer at the end of the response epoch but never get as close as at the 67 ms mark. This trend is evident in the WIM output insets above the S and T plots of Figure 4d. The WIM output in both the $0^\circ$ and $90^\circ$ directions is greater for frame 2 than it is for frame 6. Therefore the MT unit starts with a high activity at 67 ms (frame 2) but then drops to a lower value as time progresses. For motion in a $45^\circ$ direction, the time-averaged response is less than the early response (arrows and 2a and b circles in the rightmost panel of Figure 4d).

This pattern of responses across the different S, T, WIM, and MT model units accounts for the temporal “component-to-pattern” behavior of the MT unit shown in Figure 3. For the $\phi = 45^\circ$ (blue) and $\phi = 135^\circ$ (green) conditions, the $0^\circ$ direction (1a circle) is lower in the early stage of the response (67 ms) compared to the time-averaged (1b circle) response (100–200 ms). For motion of the line patterns in the $\pm 45^\circ$ directions, the situation is reversed. The responses for the $\phi = 45^\circ$ (blue) and $\phi = 135^\circ$ (green) conditions are higher at 67 ms (2a circle) compared to the time-averaged response (2b circle). The MT unit therefore peaks in the $\pm 45^\circ$ directions in the early stage but drops off over time (cf. blue and green curves for directions $\pm 45^\circ$ in top-row polar plots versus bottom row). The $\phi = 90^\circ$ case moves the bars across the MT unit at $2^\circ$/s for all test directions. This input speed matches the WIM and MT units’ speed tuning, and so the S and T energy curves are more like those in Figure 4c, and the situation depicted in Figure 4d does not arise. The $\pm 45^\circ$ direction responses are always less than the $0^\circ$ direction responses irrespective of what stage of the temporal response is examined.

Smith, Majaj, and Movshon (2005) simulations

Smith et al. (2005) have also reported on temporal effects in the direction tuning of MT neurons. They found that, for some MT pattern neurons, the early response (<100 ms) was dominated by component-like behavior, which then changed to pattern behavior in the later stages of the cell’s response profile. Superficially, this result looks very much like the Pack and Born (2001) phenomenon tested above. However, there
are some important differences. Smith et al. used plaid patterns made up of two sine wave gratings separated by 120° of orientation. The Pack and Born (2001) line stimuli differed by a maximum of 90° in orientation. Pack and Born (2001) compared the MT neuron response at 60–80 ms against the averaged response for the last 1500 ms whereas Smith et al. examined the responses in the very earliest stages after the stimulus onset (approximately 30–140 ms) with high precision (mean = 2.1 ms). The Smith et al. data are therefore able to provide insights into the very early-stage dynamics of MT neurons.

We tested one of our model MT pattern units (tuned to 0° direction and 2°/s speed) with the Smith et al. (2005) plaid stimuli (two gratings each 2 c/°, contrast = 50%, moving at 2°/s, and separated by 120°). The direction tuning of the pattern unit was measured by moving the plaid in 12 directions in 30° steps. The tuning curves for frame 2 (67 ms) and frame 5 (166 ms) are shown in Figure 5a. There is evidence for a component-to-pattern temporal effect, but it is weak and not as strong as the effect demonstrated by Smith et al. (see Figure 5a). The effect shown in Figure 5b results from a similar process to that outlined above for the Pack and Born (2001) stimuli tests; the time delay between the S and T V1 filters results in a smaller output for 0° moving plaids at the start compared to the end of the output profile. However, while examining the output traces, we discovered another possibility for the Smith et al. results.

When the plaid moves in a 60° direction (Figure 5c), one of the components moves in the 0° direction, and the other moves at 120°. This means that a WIM unit tuned to 0° in a cluster making up an MT pattern unit tuned to 0° (Figure 2a) is well activated by one component of the plaid as is a WIM unit tuned to 120°. However, this latter unit is an opponent unit (see red arrows in Figure 2b), and its activity is subtracted from the activity generated in the 0° WIM unit. Consider the case whereby there is a slight delay in the application of this inhibition. The positive activity in the 0° WIM unit would initially generate a large response in the 0° tuned MT pattern unit in response to a 60° moving plaid, but this would eventually be turned off as the inhibition from the 120° WIM unit arrives. We tested the full consequences of this “inhibition delay” by introducing a small modification to our MT model pattern units: For each WIM cluster making up a PMD, the weight (w, see Methods) controlling the activity from the opponent units (absolute β values > 90°) is made to be a function of the time (frame number). We used a capped linear function such that w = (1 − f) / 4 for f < five frames and w = −1 for f ≥ five frames.

The results of applying this inhibition delay to our model are shown in Figure 5d. Our MT model PMD starts off with typical component-like behavior. We tested the unit with a single grating as well to create the component and pattern predictions and measured the partial correlation coefficients (Smith et al., 2005). At the start of the stimulus (67 ms = frame 2), the component coefficient (Re) was 0.99, and the pattern coefficient (Rp) was 0.4, indicating that the model PMD was exhibiting a clear component-like response. As the frames progress, the pattern response becomes more dominant with the Re = −0.4 and Rp = 0.9 by frame 5. This trend mirrors that noted by Smith et al. in their population of MT pattern neurons (e.g., Figure 5a). Therefore a minor change to our model, which puts a small time lag on the influence of the WIM opponent units making up the clusters in our PMDs, enables us to mimic another form of component-to-pattern behavior observed in actual MT neurons (Smith et al., 2005).

This inhibition delay mechanism cannot explain the Pack and Born (2001) data. Their stimuli were line patterns tilted by ±45°, and so when the line segments were moving in a 45° direction across a model PMD tuned to 0°, there were no opponent units being strongly stimulated (see WIM activity insets in Figure 4d). A delay in the arrival of the inhibition would therefore have little effect on the tuning curve of the PMD. An account based on the timing differences between S and T V1 units is required to explain the Pack and Born (2001) data. From the point of view of our model MT units, these two temporal phenomena (Pack & Born, 2001; Smith et al., 2005) are quite different. The Smith et al. effect is unique to 120° plaids (with each component moving at the speed tuning value of the MT neuron) and manifests itself in the early-stage responses (30–60 ms). The Pack and Born (2001) effect is reliant on the timing difference in S and T V1 units and depends on the WIM unit properties (Figure 4). It is more apparent over longer time intervals (67–200 ms).

To account for both of these effects with a single mechanism, we considered the option of increasing the range of WIM sensor directions that provide inhibitory inputs (e.g., include the ±90° units as well). This would mean that the Pack and Born (2001) stimuli would also trigger the delayed inhibition effect and could be explained using the same mechanism used for the Smith et al. data (2005). However, there are a number of reasons for us not pursuing this option: First, the model pattern units are designed so that responses from the WIM units tuned to 0°, ±30°, ±60°, and ±90° relative to the preferred direction of the pattern unit (θ) are taken as positive evidence for motion in that direction. Activity in any of the opposite directions (±120°, ±150°, and 180°) is evidence against the overall pattern motion direction being θ and hence is subtracted from the positive activity. It would be counter to this design principle to include the activity
from the ±90° WIM units as negative evidence because they are neutral at best in terms of the actual direction of motion. Also, extending the range of inhibitory inputs would produce a model that is inconsistent with other known properties of MT neurons. In particular the Majaj et al. (2007) pseudoplaid simulations (Perrone & Krauzlis, 2008) would no longer match the data. As demonstrated in that paper (see Figure 7), the match to the MT data was very dependent upon the particular pattern of inhibition used in the model units. Therefore, we prefer to consider the WIM effect (Figure 4) and the inhibition delay as two different possible mechanisms that may underlie the component-to-pattern effect. Future studies of MT neurons that modify the Majaj et al. experiment and use plaids with a smaller component separation angle and line stimuli (à la Pack & Born, 2001) with steeper tilt angles may shed light on whether or not these two temporal effects have a common neural basis.

**“Transparency” effects**

**Stoner and Albright (1992) simulations**

Figure 6a depicts our plaid test stimuli with different intersection intensities based on those used by Stoner and Albright in 1992 (see Methods). The plaid patterns were used to test one of our model MT pattern units (tuned to 2°/s and 0°). Figure 6a shows the configuration of the plaids when they moved in the 0° direction. We also tested our MT unit with a sine wave grating moving in different directions to replicate the test carried out by Stoner and Albright (1992) to establish the direction tuning of their cells (see Methods). The result of this grating test for one of Stoner and Albright’s (1992) MT cells is shown in Figure 6b.
cell preferred motion in the 135° direction. When the plaid pattern was moved in this direction (now relabeled as 0°), the plaid with a light intersection created the largest output in the Stoner and Albright (1992) MT neuron, the dark intersection plaid generated the next largest response, and the “transparent” plaid produced the smallest response (see open square, filled circles, and filled triangles for 0° in Figure 6c). The interesting result found by Stoner and Albright (1992) is that when the plaid pattern moved in either a +67.5° or −67.5° direction relative to the PD of the neuron, then the “transparent” plaid now generated a large response and the “too dark” and “too light” plaids produced weaker responses. They inferred from this finding that the MT neuron activity changed from a pattern response to a component-like response when the plaid corresponded to a transparent stimulus with two overlaid grating patterns.

Figure 6d and e shows the comparable data set from our model MT unit. This unit was tuned to 0°, and this is verified by the direction-tuning curve for the grating stimulus (Figure 6d). The results for the different plaid stimuli (three directions and three intersection intensity values) are shown in Figure 6e. The insets in Figure 6e show cartoon depictions of typical pattern (top-black) and component (bottom-green) polar plot tuning curves to help interpret the Cartesian plots shown in Figure 6c and e. For pattern behavior, the output for a 0° direction plaid is greater than for +67.5° or −67.5° plaids, and this results in an inverted V shape in the Cartesian plots. For component behavior, the +67.5° or −67.5° directions produce a greater output than the 0° direction, and this results in a V shape in the Cartesian plots (green line in Figure 6e). Comparison of the Figure 6c and e plots shows that the model data replicate the pattern of responses found by Stoner and Albright (1992) for their MT neuron. The MT pattern unit changed from preferring plaids moving at 0° when they had dark or light intersections to preferring plaids moving at ±67.5° when the intersection was “transparent.”

To investigate why our model unit was able to emulate the Stoner and Albright (1992) MT neuron property, we examined the responses of the subunits making up the MT unit and their V1 inputs. We discovered that there was an asymmetry in the contrast along different directions of the plaid stimulus. This means that the V1 and WIM sensors are exposed to different amounts of contrast when the plaid moves at 0° compared to when it moves at ±67.5°. This is illustrated in Figure 7.

For 0° plaid motion (Figure 7a), consider a V1 spatiotemporal energy filter (either S or T) oriented vertically (see inset at top) and located over the plaid intersection at the position marked by the dotted line at A. The intensity values for the different parts of the stimulus have been marked in the figure, and X indicates the variable intensity value of the intersection (12, 76, or 136). The contrast can be approximated by comparing the intensity values along A (taking into account the relative lengths of each zone the line crosses) against those along an adjacent location (B). The same exercise can be carried out for when the plaid is moving in a ±67.5° direction (only the +67.5° is shown in Figure 7b) or when it is moving at 0°, but a V1 unit tuned to 60° is considered (Figure 7c). The approximate contrast for the two plaid directions shown in Figure 7a and b is plotted in Figure 7d. As the intensity of the intersection increases, the contrast increases for a 0° V1 spatiotemporal filter when the plaid moves in the 0° direction but decreases when the plaid moves at ±67.5°.

This theoretical trend can be verified by examining the actual output of a WIM filter tuned to 0° and 2°/s located in the region of the plaid intersection (Figure 7e) at the three levels of plaid intersection intensity used by Stoner and Albright (1992). This WIM filter uses the outputs from V1 spatiotemporal filters that are sensitive to the contrast effects shown in Figure 7d. The output increases as the intensity of the plaid intersection increases from too dark to too light (black curve). When the plaid pattern moves in a 67.5° direction (gray line) or −67.5° direction (dashed gray line), the output falls. Notice that for the “transparent” case, the output of the WIM unit is less for the 0° plaid direction than the ±67.5° directions (see dashed circles). This represents “component”-like behavior (see green arrow marked C in Figure 7e). The 0° tuned MT neuron we used for our test mainly receives its input from 0° WIM units similar to that shown in Figure 7e. Therefore, this explains why the “transparent” intersection plaid produced a smaller output in the MT unit when it moved at 0° compared to ±67.5° (see green, V-shaped, filled triangle curve in Figure 6e).

Now, because the two curves in Figure 7e cross, we can also see why the “too high” intersection case produced a pattern-like response in our MT model unit (see black downward-pointing arrow marked “P” in Figure 7e). For this case, the WIM subunit feeding into the MT unit produces a larger output for the 0° plaid direction than for the ±67.5° directions (dashed circles) and so exhibits “pattern” behavior. This explains why the MT unit ends up with an “inverted-V shaped” curve when the intersection intensity is too high (see black, open square line in Figure 6e).

Extending this logic, one might ask why then does not the “too low” condition produce strong component behavior given that for a 0° WIM unit, the 0° plaid direction produces much lower output than the ±67.5° directions (left part of Figure 7e). The answer is that the MT unit receives very little input from its primary WIM subunit (tuned to 0°) when the intersection
intensity is “too low.” Note that the black curve in Figure 7e is at zero when the intersection intensity is low. For this case, the MT unit mainly receives inputs from WIM subunits tuned to $\pm 60^o$, and we need to examine what is happening in these units to understand the “too low” case. The $\pm 60^o$ WIM units have a relatively high output under the “too low” condition, and they are exposed to the same contrast as the $0^o$ WIM units when the plaid moves at $\pm 67.5^o$ (see Figure 7b, c). Their output for the three intersection intensity values is plotted in Figure 7f. For plaid directions moving at $\pm 67.5^o$ (gray lines and bottom dashed circle), the output of a $60^o$ WIM unit is very low compared to when the plaid moves at $0^o$, and so the MT unit behaves with pattern-like responses (see blue downward arrow marked P in Figure 7f and blue line in Figure 6e).

It is actually not too surprising that our MT unit produced pattern-like behavior when the intersection intensity was low (12 units of intensity). This is the value that corresponds to a true plaid based on summation of two gratings, and our MT pattern units are designed to respond correctly to the overall motion of the plaid in this case (Methods).

**Thiele and Stoner (2003) depth-ordered plaids**

Thiele and Stoner (2003) were also able to generate component-like responses in MT pattern neurons using what they termed a “depth-ordered” plaid. This is a square wave plaid pattern in which one of the gratings has high luminance bars (121 cd/m$^2$) and the other has low luminance bars (43 cd/m$^2$). The background luminance of their plaids was 72.6 cd/m$^2$, and the intersections had a luminance of 109 cd/m$^2$. We carried out a similar analysis to that above whereby we calculated the theoretical contrast values for $0^o$ versus $\pm 67^o$ moving depth-ordered plaids. Again, there was a large contrast asymmetry whereby the contrast for the $\pm 67^o$ plaids (0.37) was higher than that for the $0^o$ direction plaid (0.1), and so using the argument we presented above for the “transparent” stimulus, we would also predict that our model MT pattern neurons will generate component-like responses for the Thiele and Stoner depth-ordered plaids.

The pattern of responses (Figure 6e) for our model MT unit across the three plaid intersection values (too low, transparent, too high) and three plaid directions ($-67.5^o$, $0^o$, $67.5^o$) is explained by the asymmetry in contrast present at the intersections of the plaid and by the fact that our MT pattern units have speed-tuned and contrast-sensitive WIM subunits as well as subunits tuned to off-axis directions. The same mechanism can explain the Thiele and Stoner (2003) result.

**Discussion**

We have been able to replicate two classes of complex behavior observed in MT neurons. The first is a shift over time from component-like responses to pattern-like responses that was first discovered by Pack and Born (2001). Using moving patterns of small line segments, they showed that initially some of their MT neurons responded to the component of motion perpendicular to the contour’s orientation. However, over a period of approximately 60 ms, the neurons signaled the true direction of the pattern independent of the orientation of the line segments. Smith et al. (2005) showed a similar effect with plaid stimuli rather than oriented line segments, but the phenomenon was the same: Some of the MT neurons stimulated with these stimuli began responding in a component-like fashion, but eventually (after approximately 60 ms), they acted as pattern cells.

We have shown that this same behavior can be replicated with our model MT pattern motion detectors (Perrone, 2004; Perrone & Krauzlis, 2008). This model uses simple integration across a small number of subunits based on our WIM model sensors (Perrone & Thiele, 2002). These WIM subunits receive their input from two types of V1-like spatiotemporal filters (S and T). The two types have different spatial and temporal properties, and one is oriented in space-time (T) whereas the other is not (S). This difference results in a delay in the temporal evolution of the S energy responses compared to the T energy. This delay, combined with the fact that the WIM unit response is a function of the difference between the S and T energy values (Equation 1, Methods), is the main reason for the component-to-pattern temporal effects we observed in our model MT units. The Pack and Born (2001) effect stems naturally from the properties of our subunits and the way the signals are combined at the WIM stage.

The addition of a slight temporal delay in the application of the inhibitory signals from our local WIM sensor clusters (Figure 2a) also enabled us to replicate data from Smith et al. (2005), which also showed a temporal change from component-like to pattern-like behavior in some MT neurons. Smith et al. suggested that their temporal effect might arise from a recurrent circuit that implements a form of divisive gain control but noted that it would need to be slower and different from the type of recurrent circuits that have been proposed at the level of V1. The mechanism we are suggesting does not rely on the divisive gain-control stage of our model (Equations 2 and 3) but does depend on a small lag being introduced at the opponency stage. The particular construction of our pattern units from WIM clusters is also important, however, and our simulations have indicated that the
Smith et al. component-to-pattern effect should only occur with plaids made up of gratings separated by more than 90°. We have previously shown how pattern behavior in our model MT units is contingent upon the inhibition created by one of the gratings making up a 120° plaid (Perrone & Krauzlis, 2008, Figure 6). We have now also shown that there may be a slight delay in the arrival of that inhibitory signal.

The second property of MT neurons that we were able to replicate also involved a shift from “component-like” behavior to “pattern-like” behavior. This effect was discovered by Stoner and Albright (1992), and they showed that some MT neurons change their preference for the component motion of plaid patterns to the overall pattern motion, depending not on the time of exposure as in the Pack and Born (2001) result, but on the “transparency” of the plaid intersections. This result maps nicely onto an effect noticed in human observers who tend to perceive either a single coherently moving “plaid pattern” or two component gratings sliding noncoherently across one another depending on the luminance of the regions of overlap (Stoner & Albright, 1992; Stoner et al., 1990). The effect has also been measured psychophysically in monkeys (Thiele & Stoner, 2003).

The apparent complexity of the behavior of the MT neurons prompted Stoner and Albright (1992, 1993) to suggest that figural aspects of the visual image unrelated to motion (such as perceptual transparency) could affect the responses of the MT neurons. They did not necessarily assume that these figural aspects were from higher levels of the visual system, and information about image properties, such as occlusion, could well come from areas prior to MT such as V2. However, the underlying assumption is that the MT neurons receive additional information unrelated to the motion signal itself. The fact that we were able to simulate this property of MT neurons using one of our basic model MT units that includes no “static” mechanisms and only relies on relatively low-level motion inputs demonstrates that there is possibly a simpler explanation for the Stoner and Albright data.

The plaid stimuli used by Stoner and Albright (1992) have an asymmetry in the amount of contrast that a motion sensor located at the grating intersections is exposed to, depending on in which direction the plaid move. For some plaid directions, the contrast increases as the intensity of the intersection increases, but for other directions, the reverse is true (Figure 7a). For the WIM subunits making up our model MT pattern neurons, this means that the output changes as the intensity of the plaid intersections changes and the plaid direction changes. “Transparent” intersections (as defined by Stoner & Albright, 1992) happen to produce less output in the MT units when the plaid is moving in the PD of the unit compared to when it is moving \( \pm 67.5° \) relative to this direction. This gives them a type of “component” behavior whereas the other two intersection intensity values (low and high) produced pattern-like behavior with plaid motion in the PD of the MT unit producing the largest output. The same underlying mechanism can also account for the Thiele and Stoner (2003) depth-ordered plaid results.

A number of explanations have been proposed for the Stoner et al. (1990) human perceptual transparency effects (e.g., Lindsey & Todd, 1996; Trueswell & Hayhoe, 1993; van den Berg & Noest, 1993), but the bulk of the proposed mechanisms include additional “nonmotion” sources of information, such as non-Fourier energy components or static image features, such as contour intersections. Because all of these mechanisms rely on additional nonmotion signals to explain the effect, they therefore require other (non-V1) inputs feeding into MT neurons in order to explain the cell behavior observed by Stoner and Albright (1992). We have shown that this is not necessary, and our model can explain both the Stoner and Albright (1992) transparency data and the Pack and Born (2001) MT temporal data.

Some researchers have considered the Stoner et al. (1990) and Stoner and Albright (1992) effect in terms of an additive component applied to a standard plaid (e.g., Movshon, Albright, Stoner, Majaj, & Smith, 2003), and it is possible to consider the Fourier components of these additive components when attempting to explain the electrophysiological or human perceptual results. Our explanation for the Stoner and Albright (1992) MT data relies on a contrast asymmetry in the plaid stimuli that arises not only from the plaid intersections, but also the alignment of the bars adjacent to the intersections (Figure 7). These local spatial factors are not easily captured in a Fourier-based analysis, and an account based on “nonadditive plaids” cannot readily explain our data.

Suggestions have been put forward for an explanation of the Stoner et al. (1990) human perceptual transparency effect that do not rely on nonmotion information, and it has been pointed out that a pointwise nonlinearity (e.g., a log transform) could account for the percept (e.g., Kim & Wilson, 1993; Stoner et al., 1990). However, these accounts cannot easily explain the MT data and do not constitute models of MT neurons in the same way as our model; they cannot explain the Pack and Born (2001) temporal effects nor many other properties of MT neurons (e.g., spatio-temporal frequency tuning).

There are other “complex” MT neuron properties that we have not yet addressed with our model but which are similar to the effects discussed above. Duncan, Albright, and Stoner (2000) showed that some MT neurons change their responses depending on the depth ordering of static regions abutting a moving
pattern in the receptive field of the neuron. The suggestion is that the neurons use contextual depth-ordering information to resolve some of the ambiguities in the motion pattern itself. Pack, Gartland, and Born (2004) showed that some MT neurons alter their responses to barber-pole stimuli depending on the shape and orientation of the aperture the stimuli were presented in. The terminators in the barber-pole patterns had a large influence on the neuron response patterns.

We have not discounted the possible role of information from outside the classical receptive field of our model PMDs and, in fact, have recently incorporated such antagonistic surrounds into our basic model (Perrone, 2012). Nor do we preclude the possibility of binocular or depth information being used to refine or modulate the PMD responses to reflect the Duncan et al. (2000) MT data. We also acknowledge that the WIM subunit signals feeding into our PMDs could be weighted differently depending on the presence of terminators to bring our model more in line with the data from MT and V1 (Pack et al., 2004). There will no doubt be other MT phenomena waiting in the wings that our basic feed-forward model cannot simulate without the addition of further external signals. What we have demonstrated in this paper is that two classes of MT neuron properties (temporal and transparency-induced component-to-pattern behavior) can arise from a basic pattern unit made up of speed-tuned WIM subunits (based on V1 inputs). We do not need additional complexity in the design of our model MT pattern units to account for these particular sets of component-to-pattern effects. This is not to say that our model will not require further modification in the future to encompass a wider range of MT phenomena.

The tests reported in this paper extend the range of MT phenomena that need to be included in a rigorous test of potential MT models. Basic tests of speed and direction tuning are no longer adequate. The model speed-tuning curves need to display the peaked (leptokurtic) shapes seen in actual MT neuron data (Lagae et al., 1993; Maunsell & Van Essen, 1983) and be able to replicate the spatiotemporal frequency maps found in MT neurons (Perrone & Thiele, 2001). The model pattern units should demonstrate the spatial effects apparent when small, separated patches of plaid components (pseudoplaids) are presented in the receptive fields of MT cells (Majaj et al., 2007) and the spatial asymmetric antagonistic surrounds found in some MT neurons (Xiao et al., 1997). As a further challenge, the models should also be able to replicate the component-to-pattern effects described in this paper.

We have now assembled an extensive catalogue of MT neuron properties that can be simulated using our image-based MT model (Perrone, 2004; Perrone & Krauzlis, 2008). The same model parameters have been used for all of these tests, and, except for the inhibitory time delay, it has not been altered for each simulation. The fact that we have been able to replicate such diverse properties with the same model indicates that the basic underlying mechanism of the model is capturing a core aspect of MT neuron processing.

Acknowledgments

Thanks to the two anonymous reviewers for their helpful comments. Supported by the Marsden Fund Council from Government funding, administered by the Royal Society of New Zealand.

Commercial relationships: none.
Corresponding author: John A. Perrone.
Email: jpnz@waikato.ac.nz.
Address: The School of Psychology, The University of Waikato, Hamilton, New Zealand.

References


