Spatial and temporal frequency selectivity of neurons in the middle temporal visual area of new world monkeys (Callithrix jacchus)

Leo L. Lui, James A. Bourne and Marcello G. P. Rosa
Department of Physiology, Monash University, Clayton, VIC 3800, Australia

Keywords: extrastriate cortex, motion selectivity, primate, receptive fields, response properties

Abstract
Information about the responses of neurons to the spatial and temporal frequencies of visual stimuli is important for understanding the types of computations being performed in different visual areas. We characterized the spatiotemporal selectivity of neurons in the middle temporal area (MT), which is deemed central for the processing of direction and speed of motion. Recordings obtained in marmoset monkeys using high-contrast sine-wave gratings as stimuli revealed that the majority of neurons had bandpass spatial and temporal frequency tuning, and that the selectivity for these parameters was largely separable. Only in about one-third of the cells was inseparable spatiotemporal tuning detected, this typically being in the form of an increase in the optimal temporal frequency as a function of increasing grating spatial frequency. However, most of these interactions were weak, and only 10% of neurons showed spatial frequency-invariant representation of speed. Cells with inseparable spatiotemporal tuning were most commonly found in the infragranular layers, raising the possibility that they form part of the feedback from MT to caudal visual areas. While spatial frequency tuning curves were approximately scale-invariant on a logarithmic scale, temporal frequency tuning curves covering different portions of the spectrum showed marked and systematic changes. Thus, MT neurons can be reasonably described as similarly built spatial frequency filters, each covering a different dynamic range. The small proportion of speed-tuned neurons, together with the laminar position of these units, are compatible with the idea that an explicit neural representation of speed emerges from computations performed in MT.

Introduction
The aim of the present study was to characterize the spatial and temporal filtering properties of neurons in the middle temporal area (MT, or V5), a key cortical field for the analysis of visual motion (Dubner & Zeki, 1971; see Britten, 2003 for review). Interest in the spatial and temporal transfer functions of neurons in different stations of the visual pathway originates from the suggestion that images can be coded and analysed by the brain in terms of different Fourier components (Campbell & Robson, 1968; Glezer et al., 1973; Maffei & Fiorentini, 1973). While at one stage this may have been seen as incompatible with feature-based representations (Hubel & Wiesel, 1962, 1968), physiological and psychophysical studies have since indicated that different Fourier channels, present at early stages of the visual pathway, can be combined in specific ways to form feature-based representations (Hess, 2003).

Here, we were interested in establishing the extent to which the neural representations of spatial and temporal frequency by cells in area MT are scale-invariant. This type of information sheds light on the type of information being extracted by neuronal populations, and allows parallels with human perception. For example, it has been reported that MT cells behave as invariant, logarithmically scaled detectors of stimulus speed: the shape of the neuronal tuning functions remain constant even though different cells have peaks corresponding to different optimal speeds (Nover et al., 2005). There is psychophysical evidence to suggest that motion detection is based on parallel spatial frequency lines, each presumably corresponding to an interconnected population of neurons analysing a corresponding region of the spectrum (Hess et al., 1998). In addition, studies of MT neurons have established a strong correlation between the preferred spatial frequency (revealed in tests using sine-wave gratings) and the preferred speed (revealed in tests using random dot fields, which comprise a number of spatial frequencies; Priebe et al., 2003). This could suggest that spatial frequency is represented as a separable parameter at the level of MT, as is the case for most cells in the primary visual area (V1; Tolhurst & Movshon, 1975; Priebe et al., 2006). In contrast, it is unclear whether temporal frequency per se is represented in the responses of MT neurons. Indeed, given that speed can be defined as a ratio between the temporal and spatial frequencies of a stimulus, an explicit neural representation of this parameter can only be achieved if a neuron’s temporal frequency tuning curves change as a function of the stimulus spatial frequency (Simoncelli & Heeger, 2001). While evidence of the existence of such ‘inseparable’ spatiotemporal tuning in MT has been presented (Perrone & Thiele, 2001), the issue of speed selectivity has remained more controversial, primarily due to methodological issues (Priebe et al., 2003, 2006; Perrone, 2006). Thus, as part of the present study we also addressed the separability of spatial and temporal frequency tuning in MT, and its relationship to the coding of speed.
Materials and methods

Animals

Single-unit recordings were obtained from MT of 14 adult New World monkeys (*Callithrix jaccus*, the common marmoset). The boundaries, topographic organization and connections of area MT have been established by earlier studies in this species (Rosa & Elston, 1998; Lui et al., 2005; Palmer & Rosa, 2006a,b). Experiments were conducted in accordance with the Australian Code of Practice for the Care and Use of Animals for Scientific Purposes, and all procedures were approved by the Monash University Animal Ethics Experimentation Committee. Some of these animals were also part of parallel anatomical investigations, having received tracer injections in the auditory cortex of the contralateral hemisphere.

Preparation

The methods for the preparation of marmosets for single-unit recordings have previously been described in detail (Bourne & Rosa, 2003). Anaesthesia was induced with ketamine (50 mg/kg; Parnell, Sydney, Australia) and xylazine (3 mg/kg; Ilum, Sydney, Australia), allowing a tracheotomy, vein cannulation and craniotomy to be performed. The dura mater overlaying the dorsal cortical surface was covered with a thin layer of silicone oil in order to prevent desiccation. After all surgical procedures were completed the animal was administered an intravenous infusion of pancuronium bromide (0.1 mg/kg/h; Organon, Sydney, Australia) combined with sufentanil (6 μg/kg/h; Janssen-Cilag, Sydney, Australia) and dexamethasone (0.4 mg/kg/h; David Bull, Melbourne, Australia), after which it was artificially ventilated with a gaseous mixture of nitrous oxide and oxygen (7 : 3). The electrocardiogram, SpO₂ level and level of cortical spontaneous activity were continuously monitored. Administration of atropine (1%) and phenylephrine hydrochloride (10%) eye drops (Sigma Pharmaceuticals, Melbourne, Australia) resulted in mydriasis and cycloplegia. Appropriate focus and protection of the corneas from desiccation were achieved by means of hard contact lenses selected by streak retinoscopy. These lenses brought into focus the surface of a computer monitor located 40 cm in front of the animal. Visual stimuli were presented to the eye contralateral to the hemisphere from which the neuronal recordings were obtained.

Electrophysiological recordings and receptive field mapping

Parylene-coated tungsten microelectrodes, with exposed tips of 10 μm (WE3001XXF; MicroProbe, Fremont, CA, USA), were directed towards area MT based on stereotaxic coordinates and sulcal morphology (Rosa & Elston, 1998). Provisional attribution of recording sites to MT during the experiment was based on mapping of multunit receptive fields, using electrodes that penetrated vertically. The dorsoventral trajectory of the electrodes resulted in a gradual shift of MT receptive field centre positions, from the lower quadrant towards the upper quadrant, and a gradual decrease in the eccentricity of receptive fields (Rosa et al., 2000). These trends allowed us to estimate the dorsal and ventral borders of MT during the recording session. Confirmation of the location of the recording sites was based on histological examination of the electrode tracks (see below). Amplification and filtering of the electrophysiological signal was achieved via a Model 1800 Microelectrode AC amplifier (AM Systems, Everett, WA, USA) and a 50 Hz eliminator (HumBug; Quest Scientific, Vancouver, Canada). The processed signal was fed into a waveform discriminator (SPS-8701; Signal Processing Systems, Adelaide, Australia), allowing the isolation of single-unit signals by means of a template-matching algorithm. The neural activity was continuously monitored by means of loud speakers, an oscilloscope (raw signal) and computer displays (processed signal, corresponding to the unit under investigation). For quantitative analyses, the spike trains processed by the SPS-8701 system were collected via a high-fidelity interface (ITC-16; Instrutech Corp., Great Neck, NY, USA) into a Macintosh computer, which also controlled the visual stimulus generation and displayed the accumulated peristimulus time histograms in real time.

The initial exploration of the receptive field boundaries was conducted using hand-held stimuli moved at various speeds, orientations and directions of motion across the screen of a 20-inch monitor (resolution 1024 × 768 pixels). Estimates of receptive field borders were then refined by presenting computer-generated moving bars or gratings while listening to the cell’s activity. Finally, in order to ensure that the stimuli used for quantitative tests were appropriately centred on the receptive fields we employed small stimuli (such as a 1° crosshair or square), flashed or moved at different points in the receptive field, in order to determine the point that elicited maximal response. Reflecting previous observations in the macaque (Raiguel et al., 1995), we found that this usually coincided with the geometrical centre of the excitable receptive field. Repeated estimations of the receptive field centre of the same cell typically yielded estimates within 1° of each other, and never exceeded 2°. This margin of error is small relative to the dimensions of the receptive fields (Lui et al., 2007).

Quantitative tests

Following the determination of the receptive field centre, neuronal response properties were studied quantitatively using computer-controlled stimuli. The stimuli for these tests consisted of rectangular patches of drifting sine-wave gratings presented against a grey background of average luminance (2.6 cd/m²). Each trial started with a 0.5-s presentation of the grey screen, during which measurements of spontaneous activity were obtained. Gratings were then presented for 2 s, with the phase varied from trial to trial, at constant speed. An intertrial interval of 4 s, during which the grey screen was presented, separated trials.

The testing of each unit started with a qualitative assessment of response selectivity, conducted with the objective of finding a stimulus configuration that was capable of eliciting robust responses. Using these estimated parameters, the cell was tested for direction of motion selectivity using gratings that drifted in directions 20° apart, allowing the determination of optimal direction by vector sum (Bourne et al., 2002); all subsequent tests employed gratings that drifted in this optimal direction. Responses to different spatial frequencies (at a constant temporal frequency of 1.8 or 2.7 Hz, selected on the basis of the qualitative test) and temporal frequencies (at the preferred spatial frequency) were then examined in sequence. In addition, a subset of neurons (94/220 cells; 42.7% of the sample) was tested with a matrix of independently varied, logarithmically scaled spatial and temporal frequencies (Perrone & Thiele, 2001; Priebé et al., 2003). Size summation tests were also conducted for most units, as reported separately (Lui et al., 2007). The protocol was then concluded with a test of contrast sensitivity and a re-test of selectivity to direction of motion, using optimal spatiotemporal parameters (Lui et al., 2005). With the exception of the contrast-sensitivity test, all other gratings had a Michelson contrast of 66%, which elicited near-maximal (average 90% of peak) responses from most cells but did not result in
supersaturation (Li & Creutzfeldt, 1984). This relatively high contrast was chosen to match empirically determined typical contrasts in natural scenes (Bex & Makous, 2002) and to address the possibility that ‘inseparable’ spatiotemporal tuning may be masked at near-threshold contrasts (Priebe et al., 2003).

**Histology**

At the end of the experiment the animal was administered an overdose of sodium pentobarbitone (Rhone-Merieux, Brisbane, Australia) and perfused transcardially with 0.9% saline, followed by 4% paraformaldehyde in 0.1 M phosphate buffer (pH 7.4). After cryoprotection by increasing concentrations of sucrose, and sectioning, alternate slides (40 μm) were stained for Nissl substance and myelin, allowing for reconstruction of electrode tracks relative to histological borders. Electrode tracks were reconstructed with the aid of small electrolytic lesions (4 μA, 10 s), which were placed at various sites during the experiment. Only cells confirmed as belonging to MT, on the basis of the pattern of myelination observed in sections stained with the method of Gallays (1979), were included in the present report. The laminar distribution of the recorded units was also determined, based on cytoarchitectural criteria illustrated by Bourne & Rosa (2006).

**Data analysis**

The responses of each cell were converted into peristimulus time histograms with a 10-ms bin width; these formed the basis of all subsequent analysis. A single trial response was computed as the mean firing rate over the entire duration of stimulus presentation. Spontaneous activity was calculated from the mean firing rate during the 500 ms before stimulus onset. The results to be reported here only include cells which responded to near-optimal stimuli at a level at least 2 SD above the mean spontaneous activity. The mean spontaneous activity was subtracted from the activity measured during the presentation of a stimulus as a preliminary step in all analyses described below. Tuning functions for different stimulus parameters were then fitted using the Matlab function ‘lsqcurvefit’ (MathWorks, Natick, MA, USA); different functions (specified in the ‘Results’ section) were used to describe the variation of different stimulus parameters. In each case the best fit for each neuron was obtained by minimizing the sum-squared error between the neuronal response and the values obtained by the function. Curve fitting was based on the entire matrix of single-trial responses rather than the mean responses to each stimulus condition. Confidence intervals for parameter estimates were computed from the Jacobian matrix and the residuals using the Matlab function ‘nlparci’. As detailed in the Results section, nonparametric statistical tests were used in most analyses due to the non-normal distributions of data.

**Results**

We have characterized the responses of 220 neurons confirmed to be in MT by histological reconstruction. The receptive fields encompassed both the upper and lower visual quadrants, and were centred on eccentricities ranging between 3 and 33° (median eccentricity 13°).

**Were the spatial and temporal tuning properties of MT neurons separable?**

Ninety-four neurons were tested with a matrix of independently varied spatial and temporal frequencies in an attempt to clarify the extent to which the responses of MT cells to these parameters are separable (Fig. 1A). This stimulus set was constructed after initial experiments which determined the approximate range of spatial and temporal frequencies to which neurons in the explored part of marmoset MT respond, so that it covered the dynamic range of the vast majority of the units studied. As summarized in Fig. 1, surfaces describing the spatiotemporal tuning (‘spectral receptive fields’; Perrone & Thiele, 2001) of separable units (i.e. those characterized by preference for stimuli of a defined optimal temporal frequency, irrespective of their spatial frequency) can be identified by isoresponse contours which are aligned with the vertical and horizontal axes (Fig. 1B). In contrast, the spectral receptive fields of inseparable units are characterized by contours which are tilted in such a way that the optimal temporal frequency depends on the spatial frequency of the stimulus. In the case illustrated in Fig. 1C, the model neuron also shows perfect tuning to the grating speed, with increments of spatial frequency being matched by a corresponding increase in temporal frequency (Perrone & Thiele, 2001). Studies in the macaque monkey area MT have shown some disagreement regarding the proportion of neurons that show true speed tuning (Perrone & Thiele, 2001; Priebe et al., 2003), so we were also interested in determining the existence of such cells in a different primate species.
Following the study of Priebe et al. (2006) we fitted a variant two-dimensional Gaussian function (Eqns 1 and 2) to the data obtained from each neuron. This model represented the data well, with a median $r^2$ of 0.89.

$$R(sf, tf) = A \times \exp\left(\frac{-\left(\log_2(sf) - \log_2(sf_{opt})\right)^2}{2\sigma_{sf}}\right) \times \left[\exp\left(\frac{-\left(\log_2(tf) - \log_2(tf_{opt})\right)^2}{2(\sigma_{tf} - \zeta(\log_2(tf) - \log_2(tf_{opt}))^2)\right)} - \exp\left(-\frac{1}{\zeta}\right)\right]$$

(1)

where

$$\log_2(tf_{opt}(sf)) = Q(\log_2(sf) - \log_2(sf_{opt})) + \log_2(tf_{opt})$$

(2)

Here $R(sf, tf)$ is the response with respect to spatial frequency ($sf$) and temporal frequency ($tf$). Free parameters were $A$, $sf_{opt}$, $tf_{opt}$, $\sigma_{sf}$, $\sigma_{tf}$, $\zeta$ and $Q$, where $A$ accounts for the size of the peak response, $sf_{opt}$ and $tf_{opt}$ relate to the position of the peak (preferred average spatial frequency and temporal frequency, respectively), the $\sigma$ parameters represent the bandwidth of the Gaussian along the spatial and temporal frequency dimensions, and $\zeta$ relates to the skew of the temporal frequency tuning curve (the relevance of this parameter is discussed below, where we present the temporal frequency selectivity). Parameter $Q$ (the exponent of power function relating the optimal temporal frequency to spatial frequency) forms the basis the present analysis. A value of $Q = 0$ denotes temporal frequency tuning that is spatial frequency-invariant (Fig. 1B) whilst $Q = 1$ denotes perfect ‘speed tuning’, where temporal frequency tuning varies in proportion to changes in the stimulus spatial frequency (Fig. 1C). Confidence interval estimates (95%) for parameter $Q$ were computed; from these we could classify cells into categories as detailed below.

The results obtained in four of the units were inconclusive, by virtue of yielding confidence intervals for $Q$ which included both 0 and 1. For three of these units the peak of the Gaussian was not adequately captured within the range of temporal frequencies tested, while one cell had a $Q$ value of 0.56 and unusually high variance. These cells were excluded from further analyses. As illustrated in Fig. 2, the majority of neurons in MT (57/90; 63.3% of the sample) were tuned to the temporal frequency of a drifting grating stimulus, with confidence intervals for parameter $Q$ overlapping 0 and excluding 1. In contrast, there were only nine speed-tuned cells (10% of the sample) for which the confidence interval for $Q$ included 1 and excluded 0. The majority of the remaining units yielded confidence intervals for $Q$ lying between 0 and 1 (22 units, 24.4% of the sample), while two units had negative $Q$ values and confidence intervals that did not extend to 0. Although these ‘unclassified’ units had inseparable spatiotemporal tuning, they showed the opposite behaviour from that expected from speed-tuned cells (they preferred progressively higher temporal frequencies as the spatial frequency of the stimulus was decreased). The distribution of $Q$ was unimodal and had a median value of 0.18. In summary, the responses of approximately two-thirds of the MT neurons were characterized by separable spatiotemporal tuning. Among the cells which did show inseparable spatiotemporal tuning, only a minority could be seen as tuned to the speed of drifting gratings.

Perrone & Thiele (2001) reached the conclusion that a majority of MT neurons have inseparable spatiotemporal tuning, including speed selectivity, based on the comparison of different versions of a two-dimensional Gaussian model fitted to the spectral receptive fields. An ‘oriented’ version of the model included a free parameter analogous to $Q$ which allowed rotation of the model around its peak, while in a nonoriented version this parameter was constrained to a value equivalent to $Q = 0$. The rationale behind this analysis is that the responses of units with inseparable spatial and temporal frequency selectivities would necessarily be better described by the oriented version of the model (e.g. Figures 3A and B), while those of units with separable selectivities would be equally well described by the nonoriented model (e.g. Figures 3E and F). For comparison, we performed a similar analysis, fitting the matrix of single-trial responses from each cell with two-dimensional log-Gaussian functions (for details, see Lui et al., 2007). We compared the residuals to the orientated and nonorientated fits using a sequential $F$-test (DeAngelis & Uka, 2003; Nover et al., 2005), where the degrees of freedom were adjusted according to the number of free parameters. According to this analysis (Fig. 4), 40 units (44.4% of the sample) yielded an $F$-value $> 3.89$ (i.e. indicating a significantly better fit by the orientated model, at the level of $P = 0.05$), while 25 units (27.8%) had an $F$-value $> 6.77$ (significantly better fit at the level of $P = 0.01$). The latter estimate is arguably more appropriate, given that the sequential $F$-test is best applied to functions that are linear in their parameters, hence justifying a stringent criterion (compare, for example, Fig. 3C and D). Moreover, there was good agreement (Table 1) between the classifications yielded by the confidence interval-based (Fig. 2) and residuals-based (Fig. 4) analyses. In Table 1, the columns indicate the category of the unit according to the confidence interval-based method (Priebe et al., 2003) while the rows indicate the classification according to the residuals-based method (Perrone & Thiele, 2001). The figures along each intersection correspond to the number of units that satisfied both criteria; for example, each of the nine units classified as ‘speed tuned’ was also deemed inseparable, while 56 of the 57 temporal frequency-tuned units were also deemed separable. In summary, irrespective of the method of analysis used, the majority of MT neurons had responses that reflected entirely, or nearly, separable spatial and temporal frequency selectivity.

Neurons with inseparable spatial and temporal frequency tuning were not homogeneously distributed across cortical layers. Considered as a group, cells which were classified as either speed-tuned or intermediate on the basis of confidence intervals of parameter $Q$ were significantly more common in infragranular layers (16/31; 51.6%)
than in the granular (6/17; 35.3%) or supragranular (9/42; 21.4%) layers ($\chi^2 = 7.20, P = 0.027$). However, the majority of the neurons classified as speed-tuned were located in layer 3 (five of nine such units).

**Selectivity for spatial frequency**

We extended the study of spatial frequency selectivity to a further 120 neurons, each of which tested at a single temporal frequency. Thus, a total sample of 214 units were used in this analysis. Typically, the relationship between neuronal responses and grating spatial frequencies could be well described by log-Gaussian functions fitted to the data (Eqn 3).

$$ R_{sf} = A \times \exp\left(-\frac{(\log_2(sf) - \log_2(sf_{opt}))^2}{2\sigma_{sf}^2}\right) $$

Here, $A$, $sf_{opt}$, and $\sigma$ were free parameters which provided estimates of the maximum firing rate above baseline, optimal spatial frequency and...
spatial frequency tuning bandwidth, respectively (see, for example, cells in Fig. 5). These functions provided accurate descriptions of the data obtained from individual units, yielding a median $r^2$ value of 0.94. In general, our sample of MT cells preferred low spatial frequencies (median 0.18 cycles/degree; Fig. 6B) and had log-scaled tuning bandwidths (represented by parameter $\sigma$) distributed around 1.11 (Fig. 6C). Cells throughout the explored range of receptive field eccentricities showed wide variation in terms of spatial frequency selectivity (Fig. 7). As a result, although the optimal spatial frequencies (Fig. 7A) were inversely correlated with eccentricity, this relationship was relatively weak ($r^2 = 0.035$, $P = 0.008$). In contrast, the bandwidth of spatial frequency tuning (Fig. 7B) showed no systematic variation with eccentricity ($r^2 = 0.01$, $P = 0.16$).

One question raised by the wide range of spatial frequency selectivities observed in MT is whether or not each unit performs as a similar spatial frequency filter, albeit optimized for a different dynamic range. For example, Fig. 6A indicates a weak ($r^2 = 0.035$), but significant ($P = 0.009$) negative correlation between optimal spatial frequency and log-scaled bandwidth, which could suggest that the representation of spatial frequency in MT deviates systematically from scale-invariance. In order to test this issue in a more rigorous manner, we applied to the spatial frequency domain an analysis analogous to that applied by Nover et al. (2005) to speed. The data from each unit were fitted with a log-Gaussian model in which parameter $\sigma$ was constrained to the median value of 1.11. The fits using this ‘fixed’ model were then compared to those in the ‘free’ model described above, in which $\sigma$ was optimized for each cell. In theory, a scale-invariant representation of spatial frequency would result in the responses of most cells being equally well described by the fixed and free models. Figure 8 summarizes the comparison of the $r^2$ values yielded by the fits to the fixed and the free log-Gaussian models.

![Fig. 4. Comparison of the $r^2$ values yielded by the orientated (free) and nonorientated (fixed) two-dimensional models fitted to the spectral receptive fields of MT neurons. The level of significance of the difference between the fits using the orientated and nonorientated models is represented according to shading of the circles. Black circles indicate cells for which the orientated model provided a better fit to the data at a level of significance of $P < 0.01$, while grey circles indicate those for which the fit was better only at a level of $P < 0.05$.](image)

Spatial frequency tuning bandwidth, respectively (see, for example, cells in Fig. 5). These functions provided accurate descriptions of the data obtained from individual units, yielding a median $r^2$ value of 0.94. In general, our sample of MT cells preferred low spatial frequencies (median 0.18 cycles/degree; Fig. 6B) and had log-scaled tuning bandwidths (represented by parameter $\sigma$) distributed around 1.11 (Fig. 6C). Cells throughout the explored range of receptive field eccentricities showed wide variation in terms of spatial frequency selectivity (Fig. 7). As a result, although the optimal spatial frequencies (Fig. 7A) were inversely correlated with eccentricity, this relationship was relatively weak ($r^2 = 0.035$, $P = 0.008$). In contrast, the bandwidth of spatial frequency tuning (Fig. 7B) showed no systematic variation with eccentricity ($r^2 = 0.01$, $P = 0.16$).

One question raised by the wide range of spatial frequency selectivities observed in MT is whether or not each unit performs as a similar spatial frequency filter, albeit optimized for a different dynamic range. For example, Fig. 6A indicates a weak ($r^2 = 0.035$), but significant ($P = 0.009$) negative correlation between optimal spatial frequency and log-scaled bandwidth, which could suggest that the representation of spatial frequency in MT deviates systematically from scale-invariance. In order to test this issue in a more rigorous manner, we applied to the spatial frequency domain an analysis analogous to that applied by Nover et al. (2005) to speed. The data from each unit were fitted with a log-Gaussian model in which parameter $\sigma$ was constrained to the median value of 1.11. The fits using this ‘fixed’ model were then compared to those in the ‘free’ model described above, in which $\sigma$ was optimized for each cell. In theory, a scale-invariant representation of spatial frequency would result in the responses of most cells being equally well described by the fixed and free models. Figure 8 summarizes the comparison of the $r^2$ values yielded by the fits to the fixed and the free log-Gaussian models.

Considering the entire sample, the free model yielded a higher median $r^2$ value.
A more reliable indication of whether the free model provided a better description of the response of individual neurons is given by the results of the sequential F-test, where the number of free parameters is taken into account. For the majority of neurons (75.7%), fixing parameter \( r \) to the median did not cause a significant decline in the quality of the fit (\( F \leq 7.40; P \geq 0.01 \); see for example the unit illustrated in Fig. 5A), implying scale-invariant tuning. Using a less stringent criterion (\( P < 0.05 \)) still resulted in only a minority (31.3%) of MT cells appearing to deviate from the expectations of a scale-invariant model (e.g. Figure 5B).

**Selectivity for temporal frequency**

We were able to characterize the temporal frequency selectivity of 180 MT neurons, after excluding a group of units that were not tested with sufficiently high temporal frequencies to fully define the shapes of their tuning curves. We found that log-Gaussian functions similar to those used to describe the spatial frequency selectivity (Eqn 3) did not provide adequate fits to the data obtained in a substantial proportion of cells. However, adding an extra parameter to allow skew provided a much better description of the data. This modified function (Eqn 4) accounted for 96% of the variance observed in over half of the sample (median \( r^2 = 0.97 \)).

\[
R(tf) = \frac{A}{C_2} \frac{\exp(-\left(\frac{-(\log_2(tf) - \log_2(tf_{opt}))^2}{2(\sigma_{tf} - \zeta(\log_2(tf) - \log_2(tf_{opt})))^2}\right) - \exp\left(-\frac{1}{\zeta}\right))}{C_0} \exp\left(-\frac{1}{\zeta}\right)
\]

Here \( R(tf) \) represents cell response with respect to varying temporal frequency \( (tf) \), while parameters \( tf_{opt}, \sigma \) and \( \zeta \) represent the optimal temporal frequency, tuning bandwidth and the skew of the tuning curve, respectively. Following this analysis, we found that the optimal temporal frequencies were distributed around a median of 2.75 Hz (Fig. 9B). All cells included in the sample displayed bandpass tuning, with 87.2% having bandwidths of \(< 2 \) units (median \( r = 1.19 \); Fig. 9C). In addition, most neurons (81.7%) displayed temporal frequency tuning that was negatively skewed (median \( \zeta = -0.22 \); see Figs 9E and 10). Significant negative correlations were found between (logarithmically scaled) optimal temporal frequency and bandwidth (\( r^2 = 0.15, P < 0.001 \); Fig. 9A) and between optimal temporal frequency and skewness (\( r^2 = 0.19, P < 0.001 \); Fig. 9D). Thus, cells preferring higher temporal frequencies also tended to be more narrowly tuned, and to have more negatively skewed tuning curves. This effect is illustrated by the tuning curves of three model neurons (calculated on the basis of the regression lines shown in Fig. 9), which represent the typical behaviour of cells with different optimal temporal frequencies, in Fig. 11. A model incorporating the systematic variation...
of parameters $\sigma$ and $\zeta$ as a function of optimal spatial frequency resulted in significantly higher $r^2$ values, as compared to a model in which these parameters were constrained to the population medians (Wilcoxon signed-rank test, $z = -2.79$, $P < 0.01$). This comparison is particularly pertinent as both models have the same number of free parameters. This indicates that the representation of temporal frequencies in MT cannot be adequately described as scale-invariant; rather, the shape of the tuning curve can be more accurately described by a model in which tuning bandwidth and skewness vary systematically as a function of the optimal temporal frequency. Finally, our results also indicate that the optimal temporal frequencies tend to increase in parallel with receptive field eccentricity ($r^2 = 0.12$, $P < 0.001$; Fig. 12A) but no corresponding variation was observed with respect to the bandwidth or skewness of the temporal frequency tuning functions (Fig. 12B and C).

Discussion

We have described the spatial and temporal frequency selectivity of neurons in area MT using first-order sine-wave gratings drifting in optimal directions. Our results indicate that, when tested with this method, most MT neurons show separable spatial and temporal frequency selectivity. Although interactions between spatial and temporal frequency selectivity were evident in approximately one-third of the units, their magnitude was in most cases incompatible with the existence of invariant speed tuning. Analysis of the data using criteria based on confidence intervals (Priebe et al., 2003) vs. comparison between a model incorporating spatiotemporal interactions and a model assuming separability (Perrone & Tiele, 2001) resulted in very similar conclusions. In conducting the latter type of analysis we accounted for the approximately logarithmic scaling of both spatial and temporal frequency tuning curves in MT (Nover et al., 2005; present results), and used a model that allowed for skewed tuning surfaces with sharp peaks (Lui et al., 2007). We believe that these modifications to the analysis technique removed some of the possible criticisms of previous studies, thereby providing accurate estimates of the spatiotemporal tuning in area MT of New World monkeys. Similar to the findings of Priebe et al. (2003) in the macaque, our data in the marmoset were normally distributed over a broad range encompassing both speed and temporal frequency tuning. However, we have in fact observed an even lower proportion of cells with inseparable spatiotemporal tuning, including speed tuning.

On one hand, our results do support the contention that some MT cells ($\sim$ 10% of the sample) are capable of spatial frequency-invariant speed tuning (Perrone & Tiele, 2001), unlike the results of most studies of V1 cells in both cats and primates (Holub & Morton-Gibson, 1981; Foster et al., 1985; Friend & Baker, 1993). On the other hand, they also indicate that the majority of neurons in MT are selective for a specific band of temporal frequencies, or at most show relatively minor deviations from the expectations of a model based on stable temporal frequency tuning. Thus, cue-invariant representation of speed must emerge from the pooling of information obtained by separate spatial frequency channels (Adelson & Bergen, 1985; Simoncelli & Heeger, 2001; Hess, 2003) and is likely to require nonlinear integration, such as the suppression of responses to certain spatiotemporal combinations, in situations where stimuli composed of multiple spatial frequencies are presented (Priebe et al., 2003, 2006).
Indeed, there is evidence to suggest that MT neurons show more explicit speed tuning when tested with broadband spatial stimuli such as bars and random dot fields (e.g. Maunsell & Van Essen, 1983; Felleman & Kaas, 1984; Priebe et al., 2003; Nover et al., 2005).

We have conducted our measurements at a relatively high level of contrast, making it unlikely that the predominance of temporal frequency tuning in our results can be explained as the result of near-threshold responses, which could in theory fail to activate the mechanism responsible for such a nonlinearity (e.g. Priebe et al., 2006). Moreover, we have focused on regions of the spatial and temporal frequency spectra corresponding to the dynamic range of the vast majority (>90%) of the units, and excluded from analysis cells for which both the rising and declining phases of the tuning curves could not be adequately captured. Thus, it is also unlikely that the existence of spatiotemporal interactions in some units may have been masked by fitting surfaces or curves to points representing somewhat less relevant portions of the units’ spatiotemporal spectra (e.g. Perrone, 2006). In addition to the improvements in the methods of analysis mentioned above, these methodological precautions may explain the lower proportions of neurons with inseparable tuning we report in the marmoset in comparison to earlier reports in the macaque. However, it is also possible that there are slight interspecific differences regarding this aspect of visual motion processing. Whether the ‘true’ proportion of speed-tuned cells in MT is closer to 10% or 25%, the basic observations remain: these cells are in the minority, and they represent one end of a continuum of response properties.

![Fig. 10. Responses of four MT neurons to variations of the stimulus temporal frequency (TF).](image)

![Fig. 11. Temporal frequency tuning curves of three hypothetical MT units, drawn to represent the mean bandwidth and skewness observed for cells with different temporal frequency optima.](image)
Fig. 12. (A) Relationship between receptive field centre eccentricity and optimal temporal frequency ($r^2 = 0.12$). There were no significant correlations between (B) bandwidth or (C) skewness and eccentricity.

Whether the inseparable spatiotemporal tuning we observed in approximately one-third of MT neurons reflects a transformation which happens at the level of MT, or is simply relayed to this area by V1 afferents, remains open to debate. Priebe et al. (2006) have reported that neurons with inseparable tuning also exist among the population of complex cells in V1, including some speed-tuned cells, and Perone (2006) suggested that an even higher proportion of such units could exist. Based on the fact that V1 and MT cells respond in an approximately similar manner (in terms of spatiotemporal separability) to stimuli formed by single sine-wave gratings, together with the notion that inputs bypassing V1 are unimportant for sustaining the responses of MT cells (Collins et al., 2003), Priebe et al. (2006) concluded that MT cells almost certainly inherit their basic spatiotemporal properties from V1. However, neurons in the koniocellular layers of the lateral geniculate nucleus, some of which are known to project directly to MT, have spatial properties that resemble those found in the magnocellular layers (White et al., 2001). Moreover, studies including the exact determination of lesioned or inactivated portions of V1 have repeatedly demonstrated that a proportion of MT neurons can respond to moving stimuli in a direction-selective manner, even in the absence of V1 inputs (Rodman et al., 1989; Girard et al., 1992; Rosa et al., 2000). Thus, neural circuits that are intrinsic to MT are able to perform relatively complex computations from inputs that show only mild direction selectivity (White et al., 2001; Schoenfeld et al., 2002). The present results, which demonstrate the predominance of separable tuning in MT and suggest that units showing inseparable spatiotemporal tuning are most commonly found in the infragranular layers, are equally compatible with the idea that inseparable tuning emerges from computations performed within MT. It is noteworthy that the infragranular layers form the bulk of the MT projections to V1 (Perkel et al., 1986; Sousa et al., 1991; Elston & Rosa, 2006), and hence the possibility that a degree of spatiotemporal inseparability exists in V1 mainly as result of feedback connections should not be discounted. Moreover, while the sample size remains too small for firm conclusions, there is preliminary evidence to suggest that the relatively few neurons which are tuned to grating speeds tend to reside in layer 3 (five out of nine such units). Thus, it is possible that speed selectivity is encoded in the output of MT to higher-order dorsal and ventral stream areas. This output would be likely to reflect the integration of multiple spatial frequency channels into a single representation which more accurately reflects the movement of ‘real world’ objects (Priebe et al., 2006).

Spatial and temporal tuning of MT cells

Logarithmically scaled Gaussian functions modelled the spatial frequency tuning curves well, and optimal spatial frequencies were normally distributed over a logarithmic scale. Over 70% of the neurons fitted the expectations of a scale-invariant model of spatial frequency representation, a proportion which closely resembles the result of the analysis performed by Nover et al. (2005) based on the speed selectivity of MT neurons. Thus, at least to a first approximation, our results are compatible with a model whereby spatial frequency is represented in a systematic manner in MT, as a result of a similar neural operation involving combinations of inputs from earlier stations of the visual pathway being performed in a modular fashion. Although such a systematic mapping is in some cases associated with columnar systems (Hubener et al., 1997; Issa et al., 2000), to our knowledge there has been no investigation of the existence of spatial frequency columns in MT. One significant deviation from a strictly scale-invariant model of spatial frequency processing is represented by the inverse correlation found between optimal spatial frequency and tuning bandwidth (Fig. 6). This correlation is also reflected in the tuning properties of V1 cells (De Valois et al., 1982), hence probably reflecting a more general limitation of the scale-invariant model when accounting for values near the extremes of the distributions (Nover et al., 2005).

A reasonable expectation is that the spatial frequencies represented by a neural population will be centred around a value determined by the sampling properties of retinal ganglion cells representing the same portion of space. However, one of the most striking characteristics of our sample is the variability of optimal spatial frequencies for cells with receptive fields located at a similar distance from the fovea (Fig. 7A; for similar observations in V1 see Schiller et al., 1976; De Valois et al., 1982). This could indicate that the information contained in the retinal afference is recombined to generate spatial filters of multiple scales covering each part of the visual field (De Valois et al., 1982). This observation is in agreement with the substantial body of psychophysical evidence indicating that motion processing is closely linked to spatial frequency channels (Hess, 2003).

Although MT neurons do resemble spatial frequency filters there are important differences in comparison to V1. In our sample the median value of parameter $\sigma$ was 1.11, which approximately equates to a half-height tuning bandwidth of 2.5 octaves. This is rather broad in comparison with results obtained in V1 of both cat and macaques, where bandwidths of 1.4 octaves are typical (Movshon et al., 1978; De Valois et al., 1982; Foster et al., 1985). Thus, MT neurons could integrate information from different spatial frequency channels, encoded in the V1 projections, in order to create a more robust representation of motion, including the speed of broadband stimuli. The responses of MT neurons are also dominated by frequencies that are lower than those observed in marmoset V1 (Lui et al., 2006), a finding that also appears to reflect observations in the macaque
cells showing strong direction selectivity, as assessed by the direction index while V1 cells included eccentricities between 6° and 25° (median 11°). Only cells showing strong direction selectivity, as assessed by the direction index (Maunsell and Van Essen, 1983) were included in the V1 sample (n = 36).

Comparison of optimal spatial frequencies of neurons in the near and mid-peripheral representations of (A) MT and (B) V1. MT neurons had receptive field centres located between 3° and 33° eccentricity (median 13°) while V1 cells included eccentricities between 6° and 25° (median 11°). Only cells showing strong direction selectivity, as assessed by the direction index (Maunsell and Van Essen, 1983) were included in the V1 sample (n = 36).

Comparisons between marmoset and macaque

Understanding common physiological features across species of different size, phylogenetic history and habits is an important requirement for the accurate interpretation of the relationship between studies in humans and animal models (Rosa & Tweedale, 2005). It is fair to expect that size differences between the macaque and marmoset will be accompanied by differences in response properties, as the decoding of information coming from the same amount of visual space (both species have frontalized eyes) has to rely on fewer neurons in the smaller animal. The brain of the marmoset is ~12 times smaller than that of the macaque (Stephan et al., 1981), a difference that is known to translate into some differences in the pattern of afferent connections to MT (Palmer & Rosa, 2006b).

The smaller eye of the marmoset somewhat limits the Nyquist frequencies despite the high densities of photoreceptors and ganglion cells (Wilder et al., 1996). Correspondingly, minimum response fields of cortical neurons centred on similar eccentricities are approximately twice as large in the marmoset (Rosa & Elston, 1998) as in the macaque (Albright & Desimone, 1987; Maunsell & Van Essen, 1987).

In the present sample, the optimal spatial frequencies of MT neurons were on average lower than those reported in the macaque, an observation that could be explained in terms of an approximately constant relationship between the resolution power of the retinal mosaic and the spatial frequency selectivity of cortical cells in the different primate species. However, it remains unclear to what extent this is really the case, given that spatial frequency selectivity varies with eccentricity and that the distribution of receptive field eccentricities has not been reported in most of the studies published to date.
Our results also reveal that cells in marmoset area MT tend to prefer lower temporal frequencies than macaque MT cells. However, if one also takes into consideration the optimal spatial frequencies for the same cells, estimates of the optimal speeds for MT neurons in the marmoset (Fig. 14) prove to be distributed over a range that is similar to that represented in macaque MT (Mansell & Van Essen, 1983). This is the case despite the peak of the distribution in the marmoset appearing to be shifted towards lower speeds (median 14.3 °/s). Also similar to what is found in the macaque, the estimated optimal speeds for MT cells vary as a function of eccentricity (Fig. 14).

Despite some quantitative differences in response properties, our results obtained in area MT of the marmoset reveal many points in common with the macaque, even though these species evolved separately for over 35 million years (Schrago & Russo, 2003). The ratio between the sizes of the macaque and human brains is similar to that relating marmoset and macaque brains. In addition, present-day New World and Old World monkeys have both developed independently from the human lineage for much of their evolutionary history (Glazko & Nei, 2003). Thus, the common physiological features revealed by the present study point to an evolutionarily conserved role of MT in the processing of visual motion which is likely to translate to the organization of the human brain.

Acknowledgements

The authors would like to acknowledge the contribution of Rowan Tweedale in correcting the text for style and grammar. Funded by a research grant from the National Health and Medical Research Council. Equipment items purchased with funds provided by the Clive and Vera Ramaciotti Foundation and by the ANZ Charitable Trust were vital for the completion of this project.

Abbreviations

A, size of the peak response; MT, middle temporal area (or V5); Q, the exponent of the power function relating the optimal temporal frequency to spatial frequency (Q = 0 denotes exclusive temporal frequency tuning whilst Q = 1 denotes perfect speed tuning); sf, spatial frequency; σtf and σsf, bandwidth of the Gaussian along the spatial and temporal frequency dimensions; tf, temporal frequency; V1, primary visual area; ζ, skew of the temporal frequency tuning curve.

References
