BRAIN RESEARCH 1368 (2011) 163–184

available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/brainres

BRAIN RESEARCH

Research Report

A metric-based analysis of the contribution of spike timing to contrast and motion direction coding by single neurons in macaque area MT

Adam J. Sachs[a,b], *Paul S. Khayat*[a], *Robert Niebergall*[a,c], *Julio C. Martinez-Trujillo*[a,*]

[a]*Cognitive Neurophysiology Laboratory, Department of Physiology, McGill University, Canada*
[b]*Division of Neurosurgery, The University of Ottawa, Canada*
[c]*Cognitive Neuroscience Laboratory, German Primate Center, Germany*

ARTICLE INFO

Article history:
Accepted 1 September 2010
Available online 8 September 2010

Keywords:
Neuron
Area MT
Motion
Contrast
Spike train
Temporal coding

ABSTRACT

Spike timing is thought to contribute to the coding of motion direction information by neurons in macaque area MT. Here, we examined whether spike timing also contributes to the coding of stimulus contrast. We applied a metric-based approach to spike trains fired by MT neurons in response to stimuli that varied in contrast, or direction. We assessed the performance of three metrics, D^{spike} and $D^{product}$ (containing spike count and timing information), and the spike count metric D^{count}. We analyzed responses elicited during the first 200 msec of stimulus presentation from 205 neurons. For both contrast and direction, the large majority of neurons showed the highest mutual information using D^{spike}, followed by $D^{product}$, and D^{count}. This was corroborated by the performance of a theoretical observer model at discriminating contrast and direction using the three metrics. Our results demonstrate that spike timing can contribute to contrast coding in MT neurons, and support previous reports of its potential contribution to direction coding. Furthermore, they suggest that a combination of spike count with periodic and non-periodic spike timing information (contained in D^{spike}, but not in $D^{product}$ and D^{count} which are insensitive to spike counts and timing respectively) provides the largest coding advantage in spike trains fired by MT neurons during contrast and direction discrimination.

1. Introduction

Single neurons in area middle temporal (MT) of macaques encode the contrast and direction of moving stimuli through variations in their mean firing rate, computed over several trial presentations of the same stimulus (Dubner and Zeki, 1971; Sclar et al., 1990). It is also known, however, that these units show a substantial variability in firing rate (spike counts) around that mean in individual trials (Buracas et al., 1998; Softky and Koch, 1993). The latter phenomenon may represent internal noise associated with the highly irregular organization and seemingly unpredictable patterns of clustering exhibited by the neurons' spike trains (Holt et al., 1996; Hu et al., 2002; Rodieck et al., 1962; Shadlen and Newsome, 1998), which ultimately affect the units' trial-by-trial coding capability.

* *Corresponding author*. McIntyre Medical Sciences Building, Department of Physiology, 3655 Promenade Sir William Osler, Room 1223, Montreal, PQ, H3G 1Y6, Canada. Fax. +1 514 398 8421.
E-mail address: julio.martinez@mcgill.ca (J.C. Martinez-Trujillo).

0006-8993/$ – see front matter.
doi:10.1016/j.brainres.2010.09.001

One possible way to reduce this internal noise, and therefore improve trial-by-trial coding of stimulus attributes by MT neurons is pooling the responses across many individual units under the assumption that each unit acts as an independent encoder of a signal embedded in noise (Green and Luce, 1975; Watson, 1990). However, adjacent neurons in cortical areas such as MT are most likely not independent, as demonstrated by measurements of noise correlation between simultaneously recorded neurons (Cohen and Newsome, 2008; Huang and Lisberger, 2009). Additionally, local field potentials (LFPs) seem to aggregate neuronal activity with a spatial extent on the order of 500 μm (Kruse and Eckhorn, 1996; Liu and Newsome, 2006), and anatomical studies have revealed complex networks of intrinsic and extrinsic interconnections among neurons (Malach et al., 1997).

A different way to improve the trial-by-trial coding capability of MT single neurons could be using the information embedded in the timing of spikes (temporal code). Previous studies using different methods have demonstrated that the timing of spikes could provide information about a stimulus direction (Bair and Koch, 1996; Buracas et al., 1998; Fellous et al., 2004; Masse and Cook, 2008; Osborne et al., 2004). Whether the same is true for coding of stimulus contrast remains unknown. Since neurons in area V1 can signal changes in stimulus contrast using spike timing information (Reich et al., 2001), MT units may possess the same capability. However, this hypothesis needs to be experimentally tested.

Here, we applied the metric-based analysis proposed by Victor and Purpura (1996) and a metric-space modification of the inner product method of Schreiber (2003) to data from two different experiments in area MT. We recorded neuronal responses in two macaques to stimuli with different contrasts and motion directions and determined the mutual information of clustering to the stimuli using the spike train metrics D^{count}, D^{spike} (Victor and Purpura, 1996, 1997), and a vector product metric, $D^{product}$ (Schreiber, 2003). We found that the temporal structure of spike trains indeed contains information about stimulus contrast and direction. Furthermore, we found that a theoretical observer model improves its contrast and motion direction discrimination when using spike timing information.

2. Results

In Experiment 1, we recorded the responses of each individual neuron to the different contrast and direction combinations of the two patterns inside the receptive field (RF). Although our main goal was to determine the contribution of spike timing to contrast coding, we have throughout the paper analyzed direction data together with contrast data. This allowed us to corroborate the results of previous studies using stimulus direction (Bair and Koch, 1996; Buracas et al., 1998; Fellous et al., 2004; Masse and Cook, 2008; Osborne et al., 2004) with a metric-based analysis, as well as to have a control for our analysis of contrast data. Fig. 1 shows responses of a typical MT unit, to the different combinations of antipreferred and

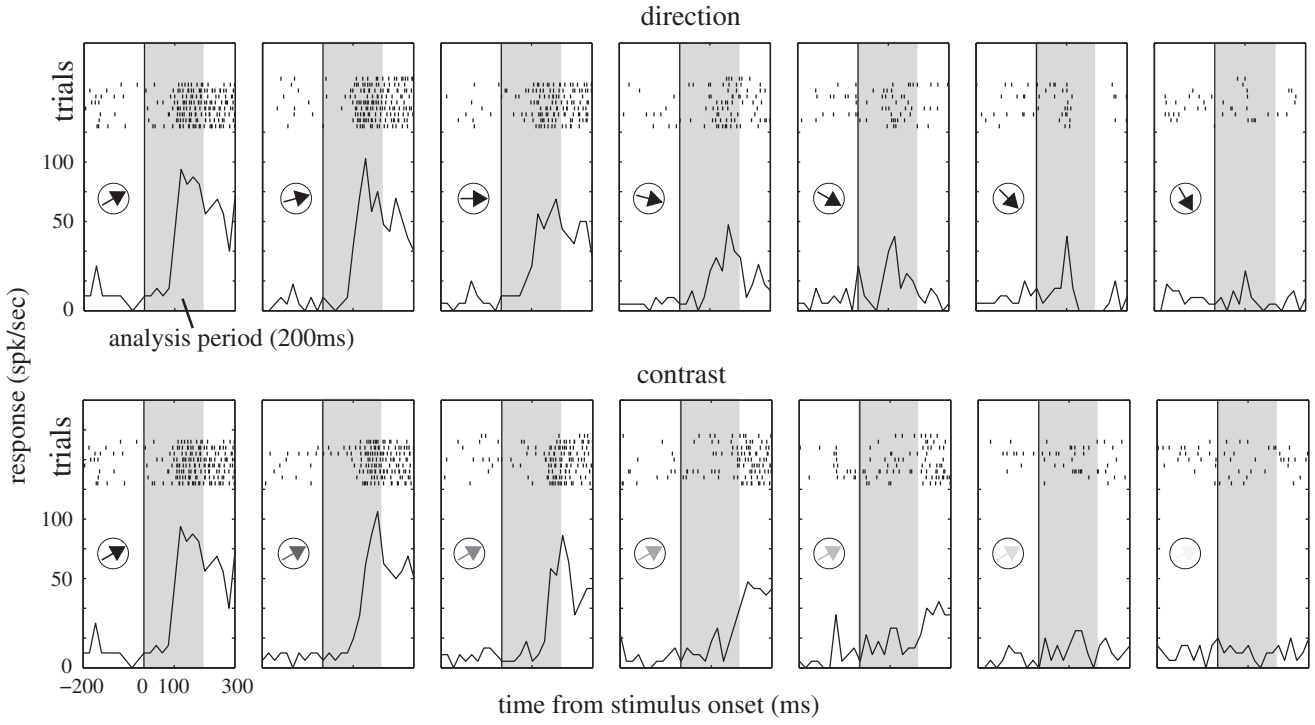

Fig. 1 – Example neuron response to different combinations of the contrast and direction configurations in Experiment 1. In each panel, the abscissa displays time from stimulus onset and the ordinate the response in spikes per second. The upper part of each panel displays raster plots of different trials used to generate the peristimulus time histograms (PSTHs) shown at the bottom. The gray shaded area represents the response time period used during the analyses. A bin width of 20 msec was used to generate the PSTHs.

test random dot patterns (RDPs). As expected from a direction-, and contrast-selective neuron, the neuron responds more strongly when the direction of the test RDP was closer to its preferred direction (direction configuration), and when that pattern had higher contrast (contrast configuration). All neurons in our sample behaved similarly; they were tuned for changes in the stimulus direction and contrast.

2.1. *Metric analysis at the single neuron level*

The metric analysis is illustrated in Fig. 2. The first row shows test RDP stimulus directions differing by steps of 15° (the antipreferred RDP direction is not shown in this figure for simplicity). The second row shows 200 msec raster plots for each stimulus direction (8 or 9 trials per stimulus in this example). This unit gave the largest response to downward motion (180°). The first step in the analysis is the selection of the stimuli to be used. We selected two stimuli subject to the restriction that the performance of a theoretical observer discriminating between them, using a metric based in firing rate alone (D^{count}), was between 55% and 82% (Fig. 2, step 1). If more than two pairs met this criterion, we chose the pair yielding the highest performance.

This procedure ensured that responses were tuned for spike counts, but that the performance of a theoretical observer using a metric based on spike counts did not saturate (i.e., that the stimuli were not easily discriminable). This is particularly important for the theoretical observer analysis when discriminating between stimuli eliciting very different firing rates, since in this scenario any metric that takes spike counts into consideration (e.g., D^{count} or D^{spike}) would perform well and likely saturate performance, rendering comparisons between them invalid. By restricting the analysis to the case in which performance is between 55% and 82%, this problem is avoided.

Following this selection, each of the eight spike trains from one stimulus category is numbered (1 to 8), and each of the spike trains from the other category is numbered (9 to 16). A D^{spike} or $D^{product}$ value is generated between all of the combinations of spike train pairs forming a 16 by 16 matrix for every value of the metric parameter (Fig. 2, step 2). Two of these matrices are displayed for different parameter values of D^{spike} and $D^{product}$. The metric values are represented by a color map in which blue represents zero, and red represents the maximum value. Theoretically, if spike trains from the same stimulus were all identical, and spike trains from different stimuli were different, this matrix would appear blue in the upper left quadrant and lower right quadrant, and red in the lower left quadrant and upper right quadrant. Each matrix is then collapsed into a single mutual information value (Fig. 2, step 3, black line), and a single value of the probability of correct response, $P_{correct}$ (Fig. 2, step 3, red line), as described in the Experimental procedures section and the Appendix A.

Continuing with this example neuron, an examination of the information entropy curve using the D^{spike} metric reveals that at $q=0$, the information estimate is 0.21 bits. As the cost of shifting a spike, q, is increased, the function achieves a maximum value of 1.0, at $q=32$, but beyond $q=64$, the mutual information decreases precipitously. Thus for this neuron the D^{spike} metric at $q=32\ sec^{-1}$ could potentially transmit about five times more information than D^{count}. Similarly, the right panel demonstrates that the information content increases for $D^{product}$ as σ is increased from 2^{-11} sec (approximately 0.5 msec) to 2^{-7} sec (approximately 8 msec) from 0.54 to 1.0. Both of these increases were significant (Wilcoxon rank-sum, $p<0.0001$). A similar trend occurs analyzing the $P_{correct}$ function (red curve using the right-sided y-axis). Analyzing for D^{spike}, at $q=0$, $P_{correct}=0.71$, and as q increases to $32\ sec^{-1}$, $P_{correct}$ increases to a maximum value of 0.87. Thus, the theoretical observer would be able to perform about 16% better at correctly assigning a spike train to its category using spike timing as with the D^{spike} metric. A similar pattern is seen with $D^{product}$, although the maximum value (at $\sigma=2^{-7}$ sec) is 0.78, somewhat lower than for D^{spike} (at $q=32\ sec^{-1}$). Nevertheless, both D^{spike} and $D^{product}$ yielded significant increases in performance (Wilcoxon rank-sum, $p<0.0001$).

2.2. *Metric analysis at the population level — Experiment 1*

We applied the analysis illustrated in Fig. 2 to each unit and then averaged across cells. In both animals, Se and Lu, there was significantly higher mutual information of stimulus clustering for both contrast and direction, when using the D^{spike} metric compared to the D^{count} metric (2-tail paired t-test, $p<0.0001$ in both cases). On the single cell level, 54 out of 102 of the cells yielded significant increases (Wilcoxon rank-sum with $p<0.05$ using the bootstrap method (see Experimental procedures section)).

In animals Se and Lu, the mean mutual information of contrast stimulus clustering was 0.26 and 0.31 bits using the D^{count} metric, 0.57 and 0.63 bits using the $D^{product}$ metric, and 0.65 and 0.70 bits using the D^{spike} metric, respectively (Fig. 3A). The differences using $D^{product}$ versus D^{count} were significant in both animals (2-tail paired t-test, $p<0.0001$ and $p=0.001$), with 53 of the neurons exhibiting significant increase (Wilcoxon rank-sum with $p<0.05$ using the bootstrap method). However, in 16 cells there was significantly higher mutual information using D^{count} than using $D^{product}$. Across the population, the differences using $D^{product}$ and D^{spike} were not significant (2-tail paired t-test, $p=0.11$, and $p=0.33$). On the single cell level, 44 out of 102 neurons yielded significantly higher information using D^{spike} over $D^{product}$, compared to 25 cells in which information was higher using $D^{product}$ (Wilcoxon rank-sum with $p<0.05$ using the bootstrap method).

In the direction trials (Fig. 3B), the mean information of stimulus clustering was 0.24 and 0.23 bits using the D^{count} metric, 0.48 and 0.63 bits using the $D^{product}$ metric, and 0.59 and 0.65 bits using the D^{spike} metric. In one animal, Se, pairwise comparisons between D^{spike} and $D^{product}$ were significant, but in the other, Lu, they were not (2-tail paired t-test, $p=0.03$, $p=0.84$). Comparisons between $D^{product}$ and D^{count} were significant (2-tail paired t-test, $p=0.00027$, $p<0.0001$). 58 cells exhibited significantly increased information using D^{spike} compared to D^{count} (Wilcoxon rank-sum with $p<0.05$ using the bootstrap method). 60 neurons exhibited significantly increased information using $D^{product}$ compared to D^{count}, however, D^{count} produced increased estimates of information in 12 neurons. Comparing D^{spike} and $D^{product}$, 39 cells yielded higher information using D^{spike}, and 23 cells

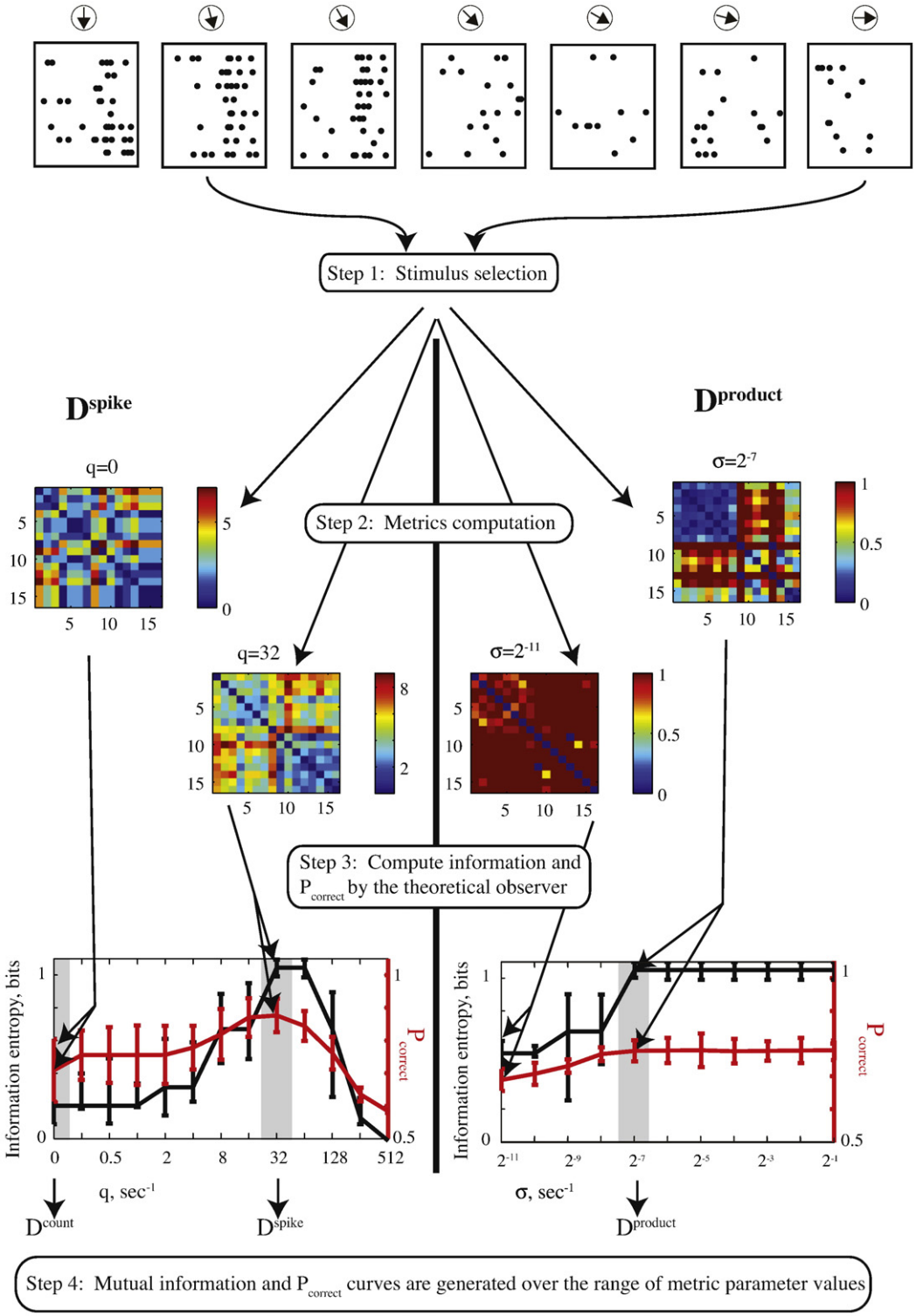

Fig. 2 – Metric and theoretical observer analysis in a single neuron example. Steps are delineated from top to bottom. First, multiple spike trains are recorded in response to different test pattern directions. Step 1 of the analysis is to select two sets of stimuli that evoke maximal $P_{correct}$ between 0.55 and 0.82 using the D^{count} metric. The corresponding sets of spike trains are then used to compute matrices (step 2 — see Experimental procedures for details) of D^{spike} values (left) and $D^{product}$ values (right) for different values of the cost parameter (left) or Gaussian kernel parameter (right). Each of these matrices is then summarized as a point (step 3) on the information curve (black) and a point in the $P_{correct}$ curve (red). Finally, the points $q=0$ on both D^{spike} curves are taken as the D^{count} values, and the maximum values are taken as the values for D^{spike} or $D^{product}$ (step 4). Error bars represent 95% confidence intervals as estimated using the bootstrap method (see Experimental procedures).

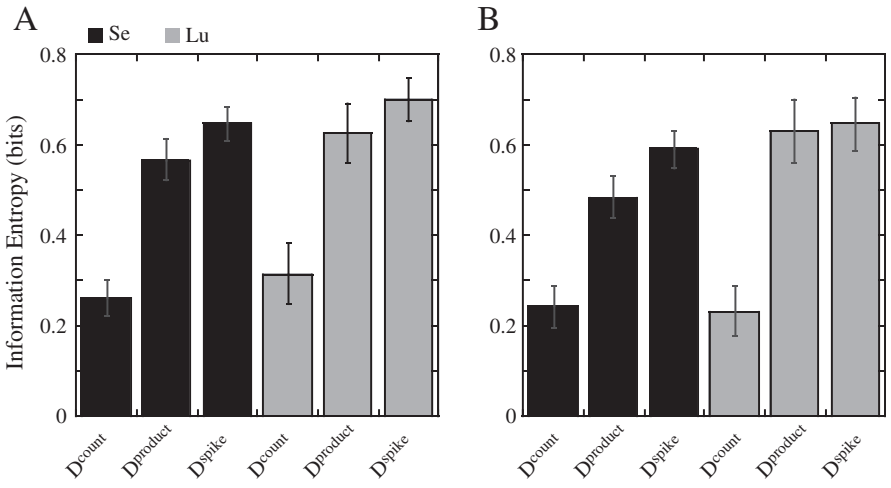

Fig. 3 – Mutual information of stimulus clustering using different metrics on spike trains recorded from area MT neurons responding to one of two contrasts of the test stimulus at the optimal direction (A), and one of two test stimulus directions at high contrast (B). Error bars represent standard errors. Black and gray bars represent data from Se and Lu respectively.

yielded higher information using $D^{product}$ (Wilcoxon rank-sum with $p<0.05$ using the bootstrap method).

We quantified the performance of a theoretical observer using an ideal discrimination strategy subject to the three choices of metrics (see Experimental procedures and Appendix A). In Se, the mean probability of a correct decision using D^{count} was $P_{correct}=0.72$, using $D^{product}$ was $P_{correct}=0.77$, and using D^{spike} was $P_{correct}=0.81$ (Fig. 4A). In Lu, these values were $P_{correct}=0.71$ for D^{count}, $P_{correct}=0.78$ for $D^{product}$, and $P_{correct}=0.82$ for D^{spike}. In both animals, there were significant differences between the performance using D^{count}, and D^{spike} (2-tail paired t-test, $p<0.0001$). Differences between the performance of $D^{product}$ and D^{spike} were significant in one animal, Se, but not the other, Lu (2-tail paired t-test, $p=0.01$, $p=0.078$). In both monkeys, the performance using $D^{product}$ was significantly better than using D^{count} (2-tail paired t-test, $p=0.0068$, and $p=0.05$).

There was a clear trend of improved performance for almost every cell using D^{spike} over D^{count} for contrast discrimination (Fig. 4B). This increase was statistically significant in 71 cells (Wilcoxon rank-sum with $p<0.05$ using the bootstrap method). Fig. 4C shows a different trend, with several neurons' spike trains yielding better performance with D^{count} than with $D^{product}$. Of these, 36 were significant. On the other hand, 55 neurons yielded improved performance using $D^{product}$. When comparing the performance of each neuron using $D^{product}$ and D^{spike} (Fig. 4D), although there is significant variability between neurons, performance was significantly better using D^{spike} over $D^{product}$ in 52 cells, compared to 37 cells in which performance was significantly better using $D^{product}$ (Wilcoxon rank-sum with $p<0.05$ using the bootstrap method). The parameters used for the $D^{product}$ (Fig. 4E) and D^{spike} (Fig. 4F) metrics varied between neurons. In the case of $D^{product}$, σ ranged from 0.2 msec (2^{-12} sec) to 250 msec (2^{-2} sec), with a mean value of 42 msec. This distribution is skewed, however, with its peak occurring at ~1 msec (2^{-10} sec). In the case of D^{spike}, the parameter q ranged from 0 sec^{-1} to 128 sec^{-1} with a population mean of 26.1 sec^{-1}. This corresponds to the majority of the distribution occurring at $q=16$ sec^{-1}, and $q=32$ sec^{-1}. Of note, 16 cells (out of 102) did not exhibit any improved performance for values of $q>0$ sec^{-1} suggesting that in this subset of cells there is no role for temporal precision coding.

For the direction discrimination data, the monkeys' mean performance values were $P_{correct}=0.70$ and $P_{correct}=0.71$ using D^{count}, $P_{correct}=0.71$ and $P_{correct}=0.74$ using $D^{product}$, and $P_{correct}=0.78$ and $P_{correct}=0.80$ using D^{spike} (Fig. 5A). For both animals there was significant pairwise improvement from D^{count} to D^{spike} (2-tail paired t-test, $p<0.0001$). In both cases there was no significant difference using $D^{product}$ versus D^{count} (2-tail paired t-test , $p=0.61$, $p=0.33$). Fig. 5B demonstrates improved performance for the majority of neurons using D^{spike} compared to D^{count}. This improvement was statistically significant in 66 cells (Wilcoxon rank-sum with $p<0.05$ using the bootstrap method). Similar to the contrast tuning case, this trend does not exist for $D^{product}$, with several cells yielding worse performance with $D^{product}$ than with D^{count} (40 cells had significant improvement using $D^{product}$, whereas 41 had significantly worse performance using $D^{product}$). Again, in general D^{spike} seems to perform the best.

The latter trend is further highlighted by a trend of improved performance using D^{spike} over $D^{product}$ (Fig. 5D). This improvement was significant in 60 cells compared to 13 cells in which performance was better using $D^{product}$ (Wilcoxon rank-sum with $p<0.05$ using the bootstrap method). Analysis of the metric parameters used in the direction discrimination task (Figs. 5E and F) yielded similar results as in the contrast discrimination task. In the case of $D^{product}$, σ ranged from 0.2 msec to 250 msec, with a mean value of 45 msec. This distribution peak occurred at approximately 4 msec (2^{-8} sec). In the case of D^{spike}, the parameter q ranged from 0 sec^{-1} to 64 sec^{-1} with a population mean of 18 sec^{-1}, and the distribution peak occurring at $q=32$ sec^{-1}. 18 cells did not exhibit any improved performance for values of $q>0$ sec^{-1}.

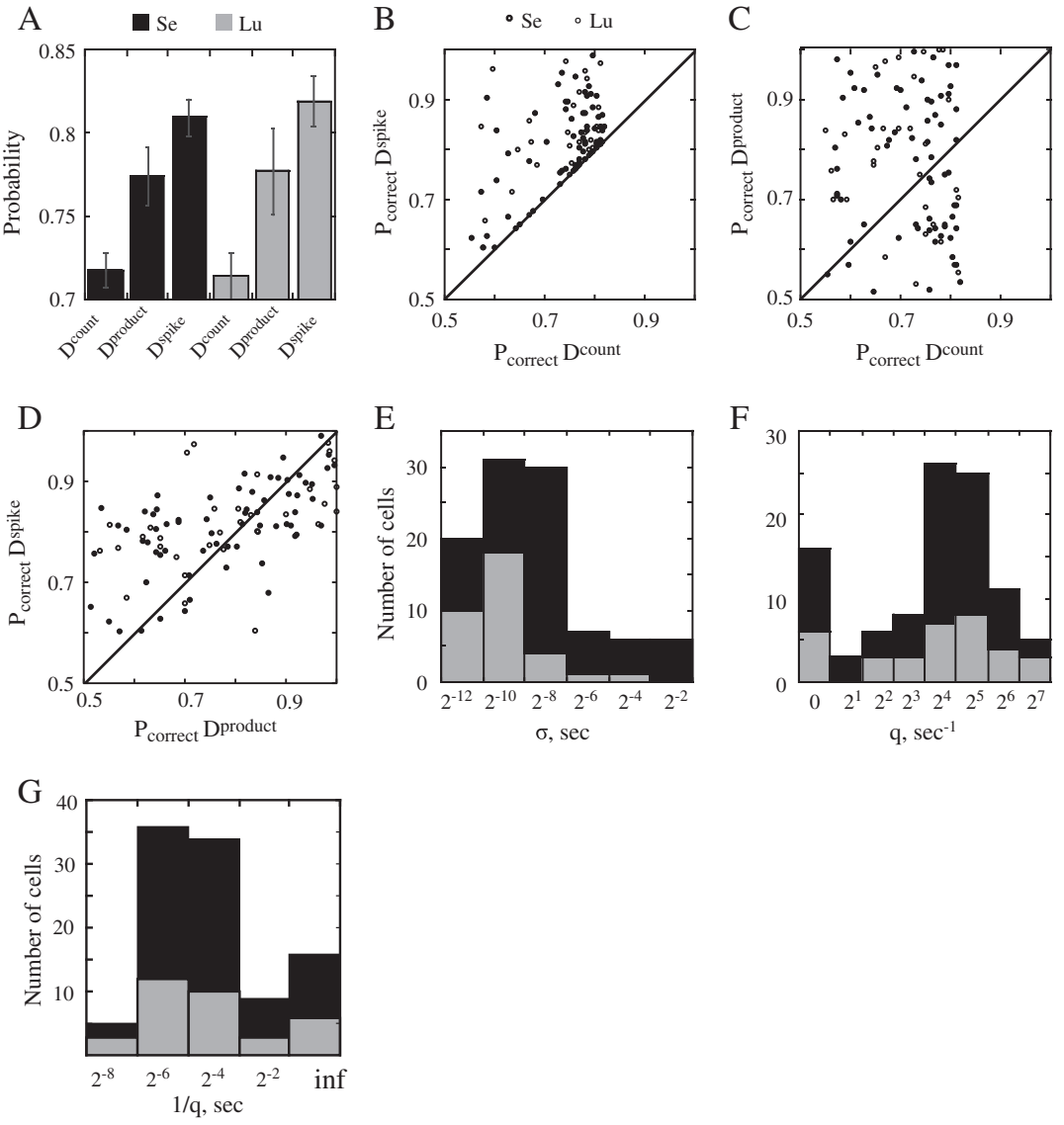

Fig. 4 – Performance of the theoretical observer model in a contrast discrimination task using D^{count}, $D^{product}$, and D^{spike} metrics (A). Joint probability graphs plotting performance of each neuron using a D^{spike} versus D^{count} (B), $D^{product}$ versus D^{count} (C), and D^{spike} versus $D^{product}$ (D). The solid line in B, C, and D is the line of equivalency. Histograms of the model parameters are shown in E for the $D^{product}$ metric (σ), in F for the D^{spike} metric (q) and its inverse is replotted in G ($1/q$).

The mean values for the cost parameter (18.0 sec^{-1} in the contrast discrimination case and 26.1 sec^{-1} in the direction discrimination case) suggests integration of temporal information occurring approximately over 40 to 60 msec. The analysis of the smoothing parameter, σ, in the $D^{product}$ metric is more difficult to interpret. Although the mean values were 42 and 45 msec, the distribution was very skewed toward much shorter estimates. There are two potential explanations for this skew. First, as Fig. 4C shows, several cells performed substantially worse using $D^{product}$ than D^{count}, and therefore the $D^{product}$ parameter may be biologically meaningless in these cases. Second, as in the single neuron case shown in Fig. 2 (bottom right), several neurons had a plateau in performance over a range of σ in the $D^{product}$ metric suggesting that the actual parameter value may not be very informative.

2.3. *Assessing metrics' behavior using surrogate data sets*

The calculation of D^{spike} and $D^{product}$ involve using an extra-parameter relative to calculating D^{count}. In order to determine if any potential bias were introduced by this extra-parameter, we recomputed D^{count} and D^{spike} on a surrogate data set constructed by replacing each spike train with an artificial spike train generated from a Poisson process with a rate parameter equal to the firing rate of the original spike train. Since this surrogate data set should contain no temporal information other than chance clustering, any improvement in performance using D^{spike} over D^{count} would be due to sample bias, and this could be used to estimate the bias in the original data. This analysis was applied to the contrast and direction configurations using the rate parameters corresponding to the data from Se and Lu separately. In the

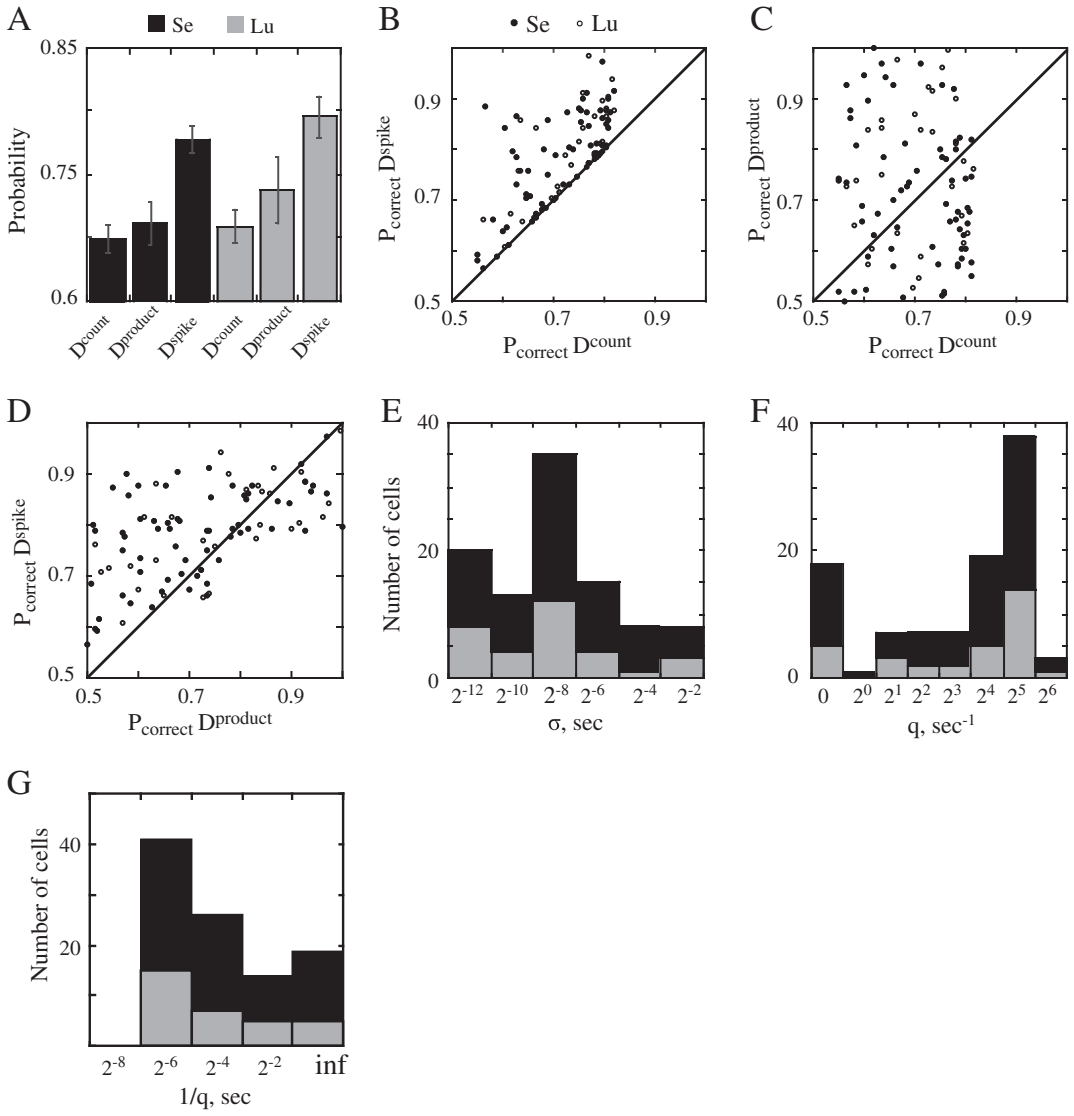

Fig. 5 – Performance of the theoretical observer model in a motion direction discrimination task using D^{count}, $D^{product}$, and D^{spike} metrics (A). Joint probability graphs plotting performance of each neuron using D^{spike} versus D^{count} (B), $D^{product}$ versus D^{count} (C), and D^{spike} versus $D^{product}$ (D). The solid line in B, C, and D is the line of equivalency. Histograms of the model parameters are shown in E for the $D^{product}$ metric (σ), in F for the D^{spike} metric (q) and its inverse is replotted in G ($1/q$).

contrast configuration (Fig. 6A), this computation yielded mean performance values of $P_{correct}=0.658$ using D^{count}, and $P_{correct}=0.663$ using D^{spike} in Se. This difference was statistically significant (2-tail paired t-test, $p=0.03$), although the mean difference was very small, representing 6% of the main effect shown in Fig. 5A. In Lu, the values were $P_{correct}=0.657$ using D^{count} and $P_{correct}=0.667$ using D^{spike}. This difference was not significant (2-tail paired t-test, $p=0.23$). In the direction configuration (Fig. 6B), the Poisson recalculation yielded values of $P_{correct}=0.647$ using D^{count} and $P_{correct}=0.653$ using D^{spike} in Se, and $P_{correct}=0.640$ using D^{count} and $P_{correct}=0.643$ using D^{spike} in Lu. These differences were not significant (2-tail paired t-test, $p=0.2$ and $p=0.5$). Thus the majority of the main effect of improved performance using D^{spike} compared to D^{count} is not attributable to bias arising from the added parameter. Furthermore, since both D^{spike} and $D^{product}$ have one free parameter the difference between them cannot be explained by differences in the number of free parameters.

Temporal correlation between spikes within a spike train have been previously proposed as a means of robust signaling, overcoming synaptic noise (Salinas and Sejnowski, 2001). In order to assess the contribution that temporally correlated spikes within a spike train might have on temporal coding, we generated a second set of surrogate data for re-analysis. For every spike train, each spike time was randomly reassigned to another spike train *within* the same stimulus category. This method maintains the spike count for every spike train (and consequently D^{count} is unaffected), and maintains spike times for every category, but eliminates the effect of temporal correlation between spikes within a given trial. We found

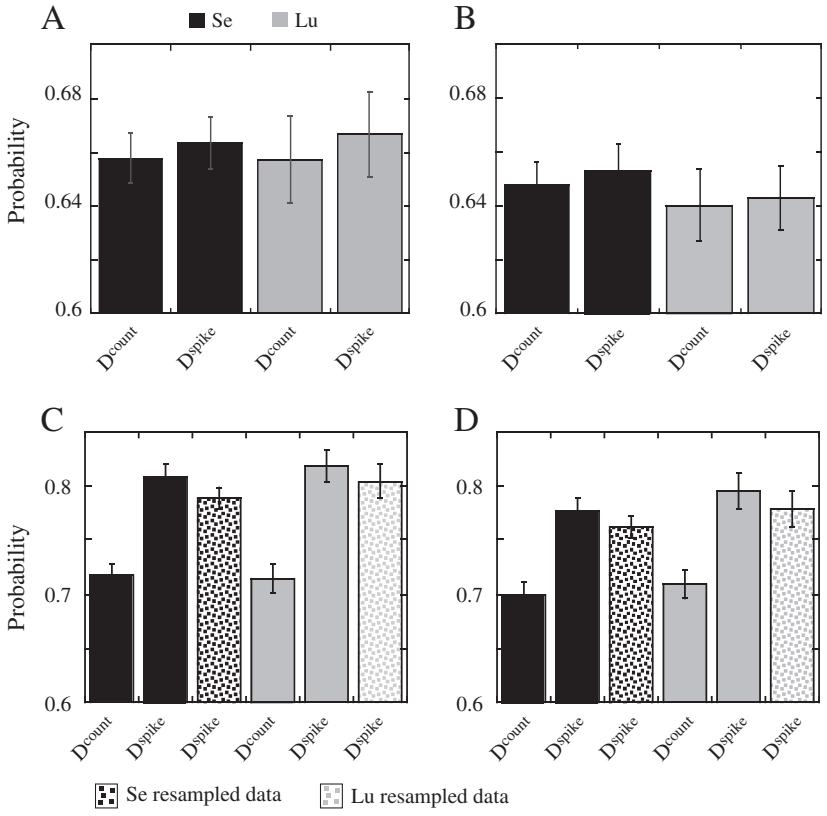

Fig. 6 – Theoretical observer performance using D^{spike} and D^{count} on surrogate data sets constructed by replacing the spike train generated from each trial (from Experiment 1) with a Poisson process in the contrast categories (A) and direction categories (B). Theoretical observer performance using D^{spike} and D^{count} on real (as in Figs. 5 and 6) and surrogate data sets constructed by randomly exchanging spike times between trials of the same category in contrast conditions (C) and direction conditions (D).

that the difference in population means between original and re-sampled data was small (~0.015) and not significant (2-tail paired t-test, contrast: $p=0.18$, $p=0.51$; direction: $p=0.31$, $p=0.46$) (Fig. 6). Therefore, the contribution of temporally correlated spikes does not appear to be the main source of usable spike timing. Furthermore, since this recalculation effectively destroys the original sequences of interspike intervals, potential information stored in these intervals is unlikely the source of the spike timing information.

2.4. *Effect of the analyzed response period on the metrics behavior*

All of the preceding analyses have been done using the first 200 msec of response following stimulus presentation. It may be that this has an influence on the contribution of spike timing to direction and contrast coding and that such a contribution becomes smaller during later periods of the response. In order to test this hypothesis, we reapplied the

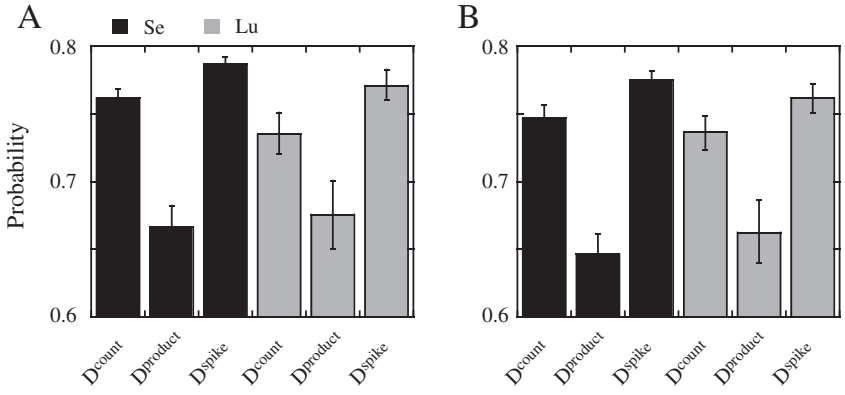

Fig. 7 – Theoretical observer performance using sustained response of MT neurons to the stimulus (350–550 msec after stimulus presentation) for contrast discrimination (A) and motion direction discrimination (B).

analysis to later response periods also using a 200 msec interval, but starting 350 msec after the stimulus onset. The trend of improved performance using D^{spike} over D^{count} exists over this time interval (2-tail paired t-test, contrast (Fig. 7A): $p=0.005$, $p=0.0016$, direction (Fig. 7B): $p=0.0026$, $p=0.005$), though this effect was significantly smaller than the one using the initial 200 msec. This occurred both because of improved performance using D^{count}, and of reduced performance using D^{spike}. The clearest difference between analyses over this interval and over the initial 200 msec was the substantial and significant reduction in performance using the $D^{product}$ metric. This occurred in the data from both monkeys, and in both conditions (2-tail paired t-test, contrast: $p<0.0001$, $p=0.0030$, direction: $p<0.0001$, $p=0.013$).

In summary, our choice of analysis time window does have an effect on the behavior of the different metrics. The performance of D^{count} improved, the one of D^{spike} slightly decreased and the one of $D^{product}$ suffered a major drop. Thus, although somewhat reduced relative to early response periods, spike timing (quantified by D^{spike}) seems to provide a coding advantage during the later sustained response period. The fact that the differences between D^{spike} over D^{count} were much larger than the ones demonstrated in Figs. 6A and B using the surrogate data sets suggests that the superior performance of the former metric was not due to the additional free parameter.

2.5. *Population results — Experiment 2*

In Experiment 1, we used a configuration with pairs of stimuli in the RF. One may argue that this has played a role in our results and they cannot be generalized to other situations, for example when a single stimulus is inside the RF. In order to test this possibility, we repeated the analysis of mutual information of stimulus clustering and theoretical observer performance on the data set collected in Experiment 2. Here, only one RDP moving in different directions (spaced every 30°) was presented inside the cell's RF (see Experimental procedures, Task). Using the D^{count} metric, the mean information values in Se and Lu were 0.40 and 0.25, respectively (Fig. 8A). When D^{spike} was used, the mean information values increased to 0.72 and 0.83. In both animals, the differences were significant (2-tail paired t-test: $p<0.001$ in both cases, with significant increases in 45 cells using Wilcoxon rank-sum, $p<0.05$ with the bootstrap method).

Using $D^{product}$, in both Se and Lu, the mutual information was 0.62. This value was significantly larger than when using D^{count} (2-tail paired t-test, $p<0.001$ in both animals), and it was significantly smaller than using D^{spike} in Lu (2-tail paired t-test, $p<0.001$), but it was not significant in Se (2-tail paired t-test, $p=0.13$). At the single unit level, 38 cells (out of 103) yielded higher mutual information using $D^{product}$ over D^{count}, compared to 14 cells in which mutual information was higher using D^{count} over $D^{product}$. 48 cells yielded

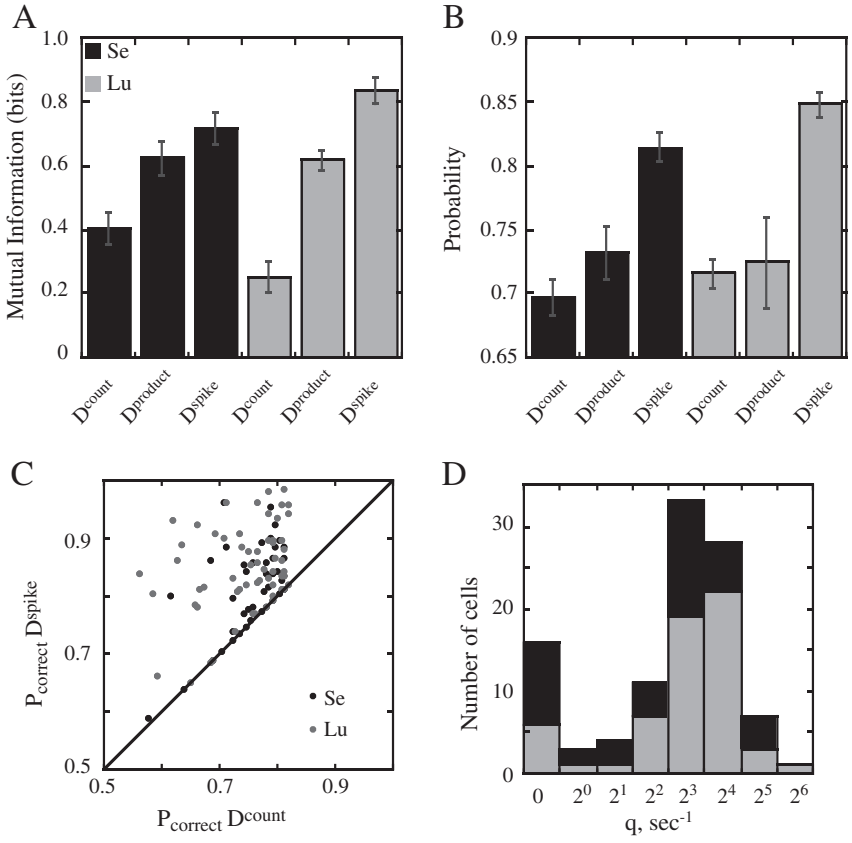

Fig. 8 – Results of Experiment 2. (A) Mutual information of stimulus clustering using the D^{count}, and D^{spike} metrics. (B) Theoretical observer performance in a direction discrimination task using the D^{count}, and D^{spike} metrics. (C) Joint probability graphs plotting performance of each neuron using a D^{spike} versus D^{count}. (D) Histogram of the cost parameters used with the D^{spike} metric.

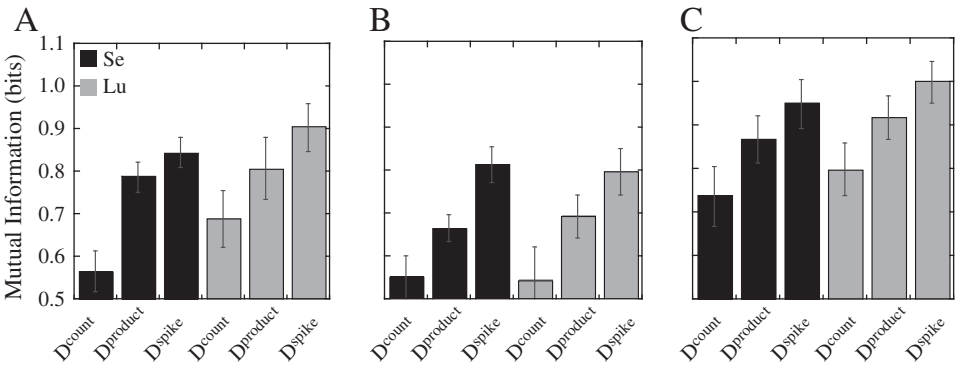

Fig. 9 – Mutual information of stimulus clustering using D^{spike}, $D^{product}$, and D^{count} over all conditions in the contrast configuration of Experiment 1 (A), the direction configuration of Experiment 1 (B), and in Experiment 2 (C). The abscissa displays the metric and the ordinate the mutual information in bits. The colors represent data from different animals.

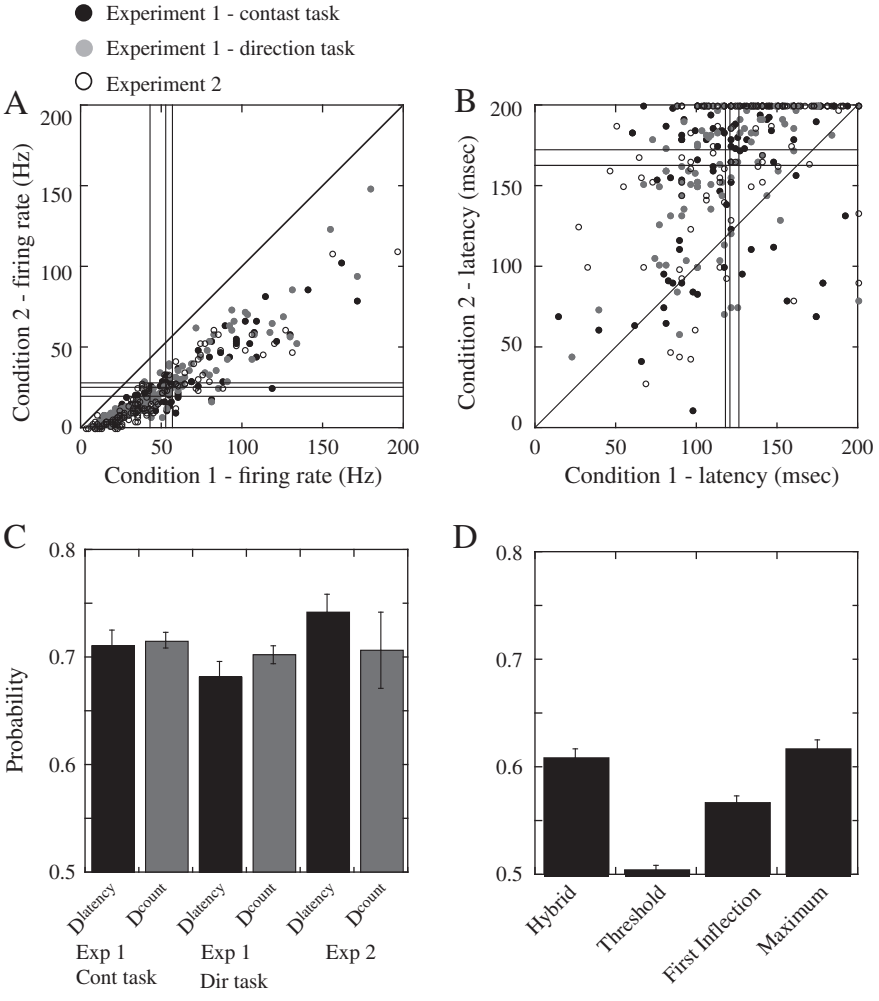

Fig. 10 – Theoretical observer performance using D^{count} and $D^{latency}$. In this figure, neurons from the two monkeys are pooled. A: mean firing rate for each neuron at condition 1 and condition 2 (see Results) in contrast and direction configuration of Experiment 1, and Experiment 2. B plots the mean latency for each neuron at condition 1 and condition 2 in the contrast and direction configurations of Experiment 1, and Experiment 2. In A and B, the gray horizontal and vertical bars represent the mean values across neurons in the three tasks. Theoretical performance in the three tasks using $D^{latency}$ compared with D^{count} (C). Theoretical performance in the contrast task of Experiment 1 using $D^{latency}$ in which latency is computed using 4 different methods.

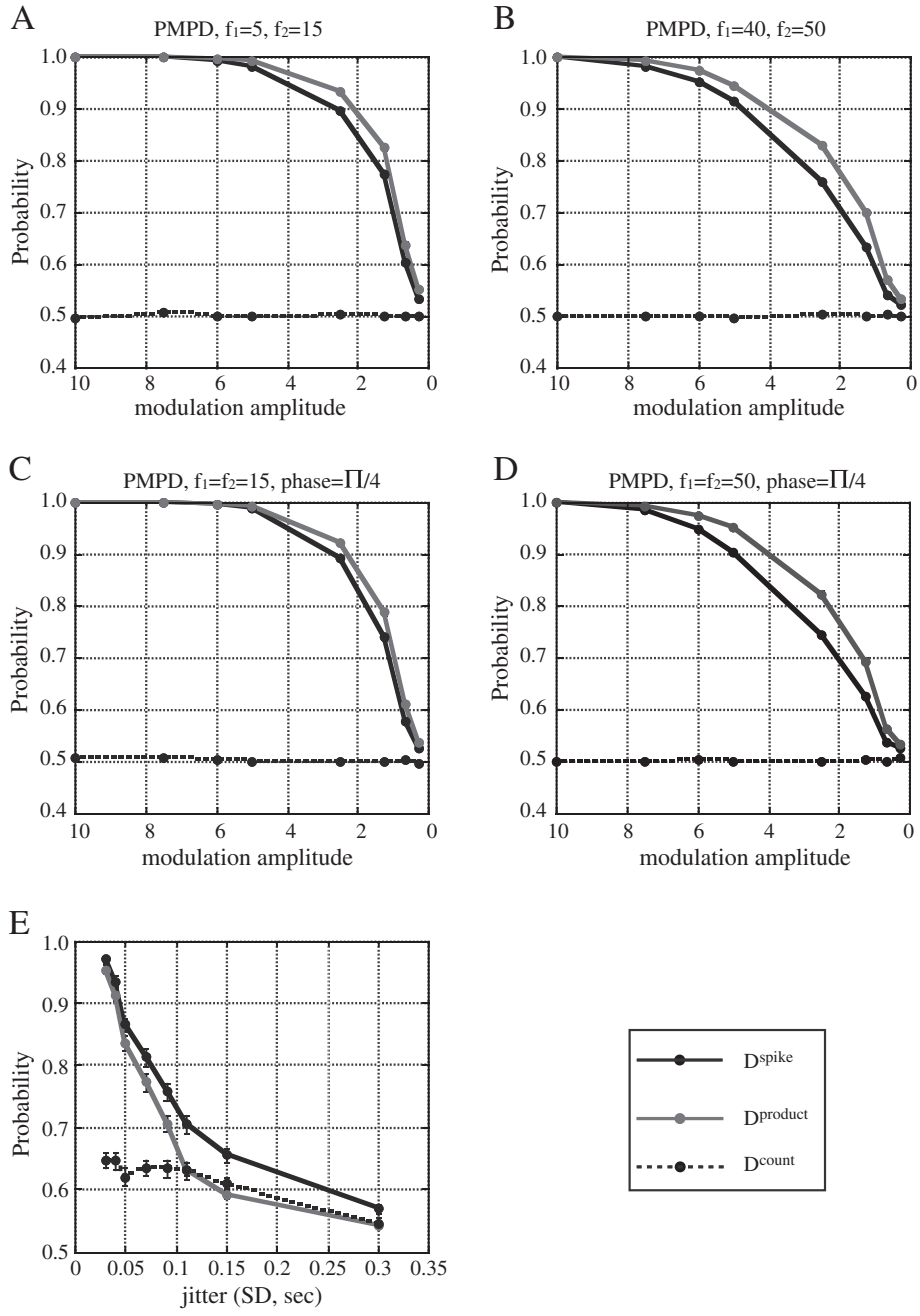

Fig. 11 – Theoretical observer performance using the D^{spike}, $D^{product}$, and D^{count} metrics for simulated spike trains generated by the PMPD model with zero phase-offset and carrier frequency parameters $f_1 = 5$ Hz, $f_2 = 15$ Hz (A), carrier frequency parameters $f_1 = 40$ Hz, $f_2 = 50$ Hz (B), between-category phase-offset of Π/4 and carrier frequency $f_1 = f_2 = 15$ Hz (C), between-category phase-offset of Π/4 and carrier frequency $f_1 = f_2 = 50$ Hz (D), and the spatial jitter model (E).

higher information using D^{spike} over $D^{product}$ compared to 11 cells with higher mutual information using $D^{product}$ over D^{spike} (Wilcoxon rank-sum with $p<0.05$ using the bootstrap method).

Using D^{count}, mean performance values of the theoretical observer model were $P_{correct}=0.70$ and $P_{correct}=0.72$ in Se and Lu (Fig. 8B). Using D^{spike} improved performance to $P_{correct}=0.81$ and $P_{correct}=0.85$. The differences were significant (2-tail paired t-test: $p<0.001$ in both cases with significant increases in 56 cells using Wilcoxon rank-sum, $p<0.05$ with the bootstrap method). Using $D^{product}$, $P_{correct}$ were 0.73 and 0.72 in Se and Lu respectively. These values were not statistically different than using D^{count} (2-tail paired t-test, $p=0.17$ and $p=0.9$), and were statistically lower than using D^{spike} (2-tail paired t-test, $p<0.001$ in both animals). The single unit analysis demonstrated improved theoretical observer performance in 24 cells using $D^{product}$ over D^{count} compared with 49 cells in which performance was better using D^{count} (Wilcoxon rank-sum with $p<0.05$ using the bootstrap method). 60 cells produced higher

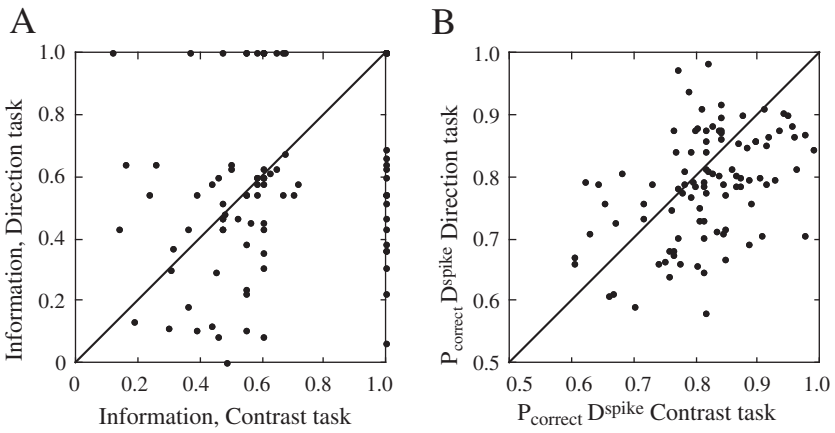

Fig. 12 – Effect of task. Information content (A) and theoretical observer performance (B) for the direction task versus the contrast task of Experiment 1 for each neuron.

levels of performance using D^{spike} over $D^{product}$ compared to 16 cells in which performance was better using $D^{product}$ (Wilcoxon rank-sum with $p<0.05$ using the bootstrap method).

We found that for most neurons, there was a clear improvement in performance using D^{spike} over D^{count} (Fig. 8C). The majority of neurons had their ideal performance optimized at cost values of $q=8$ sec^{-1} and $q=16$ sec^{-1} (Fig. 8D). The results using $D^{product}$ were very similar to the ones in Fig. 4. For most cells, the performance of this metric was in between D^{count} and D^{spike}. Thus, the results of the metric analysis using data from Experiment 2 are similar to the results of Experiment 1. Using D^{spike} provides the largest coding advantage followed by $D^{product}$ and finally D^{count}.

2.6. *Mutual information for all stimuli*

One possibility is that our selection of two stimuli has somewhat influenced the metric behavior. In order to control for this variable we quantified the mutual information of stimulus clustering comparing the metrics D^{count}, $D^{product}$, and D^{spike} using all 7 contrast and direction conditions of Experiment 1 (Figs. 9A and B, respectively) and the 7 conditions (motion directions separated by 30° intervals) of Experiment 2 (Fig. 9C) to allow comparison between experiments.

In the contrast configuration of Experiment 1 (A), using the D^{count} metric, the mean information values in Se and Lu were 0.57 and 0.69. When D^{spike} was used, the mean information values increased to 0.84 and 0.90 respectively (2-tail paired t-test, $p<0.001$ in both cases). This difference was significant in 37 cells (Wilcoxon rank-sum, $p<0.05$ using the bootstrap method). Using $D^{product}$, mean information values were 0.78 and 0.82 respectively. The first of these values was significantly higher than using D^{count}, but in the second monkey this difference was not significant (2-tail paired t-test, $p=.0001$, and $p=0.15$). These values did not achieve a significant difference from the D^{spike} case (2-tail paired t-test, $p=.14$, and $p=0.14$). In the direction configuration of Experiment 1 (B), using the D^{count} metric, the mean information values in Se and Lu were 0.55 and 0.54. When D^{spike} was used, the mean information values increased to 0.81 and 0.80 respectively (2-tail paired t-test: $p<0.001$ in both cases, significant increases in 44 cells using Wilcoxon rank-sum, $p<0.05$ with the bootstrap method). Using $D^{product}$, mean information values were 0.66 and 0.70 respectively. These values were significantly higher than using D^{count} (2-tail paired t-test, $p=0.03$, and $p=0.05$) and less than using D^{spike} (2-tail paired t-test, $p<0.0001$, and $p=0.02$). This trend was also seen in Experiment 2 (C), in which using the D^{count} metric, the mean information values in Se and Lu were 0.74 and 0.80. When D^{spike} was used, the mean information values increased to 0.95 and 1.00 respectively (2-tail paired t-test, $p<0.001$ in both cases, significant in 28 cells using Wilcoxon rank-sum, $p<0.05$ with the bootstrap method). Using $D^{product}$, mean information values were again intermediate between D^{count} and D^{spike}: 0.89 and 0.90 respectively (significant compared with D^{count}: $p=0.023$ and $p=0.05$, but not achieving a significant difference compared with D^{spike}: $p=0.06$ and $p=0.08$). These results demonstrate that our selection of stimuli cannot account for differences in information content between the different metrics.

2.7. *Bias estimation*

We estimated the bias in the mutual information calculations and subtracted it to produce final bias-corrected estimates (see Experimental procedures — Metric-based data analysis for details of bias computation). In experiment 1, using the D^{count} metric, the estimated bias was 0.03 ± 0.007 SE (standard error) for the contrast, and 0.04 ± 0.007 SE for the direction configuration. Using D^{spike}, the bias estimates were 0.1 ± 0.007 and 0.12 ± 0.008 in the contrast and direction configurations. Using $D^{product}$, the bias estimates were 0.02 ± 0.001 SE in both contrast and direction configurations. In Experiment 2, the estimated bias was 0.13 ± 0.001 SE using the D^{count} metric, 0.03 ± 0.001 SE using D^{spike}, and 0.05 ± 0.001 SE using $D^{product}$. These estimates were very small relative to the differences in mutual information between the different metrics and thus, cannot be the cause of such differences.

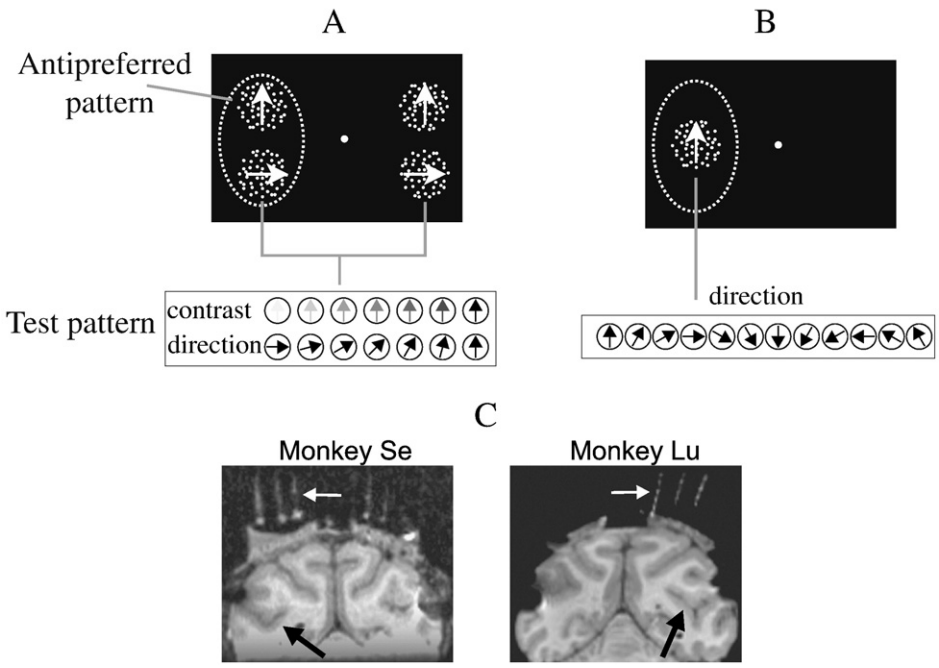

Fig. 13 – Stimulus configuration. In Experiment 1 (A), a test and a null (antipreferred) RDP were positioned inside the RF (dashed circle) and a similar pair was positioned in the opposite hemifield. In the direction configuration, the direction of motion of the test RDP was varied in steps of 15° (bottom panel upper row). In the contrast configuration, the contrast of the test RDP was varied (bottom panel lower row). In Experiment 2 (B), only one RDP was positioned inside the neuron's RF. The direction of motion was varied in steps of 30°. In both experiments the animal detected a change in the contrast of the central fixation spot. (C) Magnetic resonance brain images of the animals Se and Lu. The dark arrows illustrate oil-filled capillaries indicating the orientation of the electrode penetrations. The walls of the recording chamber were 5 mm away from the outer capillary. The white arrows indicate the anatomical location of the area targeted during the penetrations.

2.8. *Analysis of response latencies*

From the previous result, we may conclude that aspects of spike timing quantified by D^{spike} provide a coding advantage relative to D^{count} and $D^{product}$, and that this result is not due to the existence of an additional free parameter when computing this metric or the correlated timing of spikes within a train. We will discuss the differences with $D^{product}$ later, and now concentrate on D^{spike}.

This metric is computed by taking into account the number of spikes contained in two spike trains as well as their position in time. Although the metric does not assign different weights to spikes occurring at different times from stimulus onset, it is possible that some of the spikes are more informative than others and contributed to most of the advantage provided by this metric. For example, previous studies have shown that the latency of activity onset conveys information about a stimulus (Raiguel et al., 1999). We reasoned that since latency relies on the time of the first spike or spike burst, the D^{spike} metric undoubtedly captures this feature. For example, two spike trains in which the first spike or burst occurs at the same time will lead to a low value of D^{spike} relative to spike trains in which they occur at different times since the only difference would be the cost of shifting these initial spikes in time. Therefore, one could hypothesize that the timing of these first spikes that determine latency is the only feature that makes D^{spike} the best performing metric. If this were the case, comparing the performance using latency against the one using firing rate should be similar to comparing D^{spike} to D^{count}.

To begin this analysis we first compared the mean firing rate across trials in the preferred stimulus condition to that of the non-preferred condition (Fig. 10A), followed by the mean latency of activity onset in the preferred versus non-preferred condition. Finally we evaluated a metric based on latency and compared its performance to D^{count}. As expected, the mean firing rate in the preferred stimulus condition was lower than the one in the non-preferred condition (Fig. 10A). We plotted the mean latency of activity onset (across several trials) in the non-preferred condition against the mean latency of activity onset in the preferred condition. The majority of points in all three cases fall above the unity line, although the data were more scattered than for the mean firing rate (Fig. 10B). This suggests that in general our estimate of mean latency was consistently lower in the preferred relative to the non-preferred condition. However, mean firing rate and latency measurements do not provide an estimate of these parameters' variability in individual trials. Since we were interested in a measurement of performance that would incorporate such variability, we subjected these two measurements to separate signal detection analyses and compared their results.

To be consistent when comparing the performance using latency against the one using spike counts, we computed $D^{latency}$ — the difference between latency of onset between

pairs of spike trains (see Experimental procedures, Latency analysis). In all three cases (Fig. 10C), the theoretical observer performed substantially better than chance using either D^{count} or $D^{latency}$ ($p<0.05$, Wilcoxon test for all bars relative to the 50% performance line; note that data were pooled across both monkeys), indicating that both firing rate and latency provide information about the stimulus. Furthermore, the observer performance was similar using $D^{latency}$ compared to D^{count} ($p>0.05$, paired t-tests). We also recomputed the latency using different methods (see Experimental procedures) and found that the observer's performance did not improve significantly, in fact in most cases is decreased (Fig. 10D).

In summary, the results show that $D^{latency}$ performance is similar to D^{count}. Thus, although $D^{latency}$ may provide information about the stimulus and could potentially contribute to performance, it is less informative than D^{spike}. This suggests that latency cannot be the sole feature providing benefit to the D^{spike} metric over D^{count}. One possibility is that since the D^{spike} metric is sensitive to both of these spike train features, the associated advantage is due to both features' composite effect.

2.9. *Assessing metrics behavior using simulated spike trains*

In order to further explore properties of the spike trains that may underlie the previous results, and to better understand the differences between D^{spike} and $D^{product}$, spike trains were artificially generated (as described in the Experimental procedures section), and the performance of the theoretical observer was computed. Fig. 11 plots the results of 100 simulations (in each subplot), each with 15 trials of 200 msec generated for each category. The mean performance is plotted with error bars representing standard errors. Note that in Figs. 11A–D the error bars are smaller than the data points.

Fig. 11A shows the performance (see Experimental procedures) at distinguishing spike trains generated by the periodically modulated Poisson distribution (PMPD) model with frequency parameter $f_1=5$ Hz from those generated by the PMPD model with frequency parameter $f_2=15$ Hz. No phase-offset was used in this simulation. The D^{spike} and $D^{product}$ metric performed much better than D^{count} over the entire tested range of modulation amplitude, m. Over this range, the $D^{product}$ outperformed the D^{spike} metric. When PMPD spike trains were generated with higher frequencies, $f_1=40$ Hz and $f_2=50$ Hz the relationship between metric performance was the same. Over the range of modulation amplitudes, $D^{product}$ yielded better performance than D^{spike}. To assess how phase-offset might affect each metric, we repeated the simulation with the same carrier frequency (15 Hz in Fig. 11C, 50 Hz in Fig. 11D) in both categories, but with a phase-offset of $\Pi/4$. In both cases the relationship between the metrics' performance remained the same: Over the range of modulation amplitudes $D^{product}$ yielded better performance than D^{spike}.

Finally, we used the temporal jitter of individual spikes as a parameter (instead of modulation amplitude). In this scenario, the metric performance was very different (Fig. 11E). Over the tested range of the spike jitter (σ), D^{spike} performed significantly better than $D^{product}$, while D^{count} performed generally the worst, although the performance of these latter metrics seems to overlap over a range of spike jitters. These results resemble the ones described in the different data figures of both experiments.

In summary, the results of these simulations and analysis suggest that periodicity within the spike train is better accounted for by $D^{product}$, while the temporal non-periodic structure introduced by spike jitter was better accounted for by D^{spike}. Since in our neural data the latter metric led to the highest information content and performance, one explanation may be that the recorded spike trains more closely resemble the simulated data with spike jitter, containing non-periodic time information.

2.10. *Stimulus effect*

Is the information content of MT neurons stimulus-dependent? To address this question, we plotted information using D^{spike} in the direction task versus the contrast task of experiment one for each neuron. If the values were tightly clustered around the diagonal, it would imply that the information coding is independent of stimulus. Fig. 12 shows that this is not the case: any given neuron may exhibit high information coding in one task, but low on the other (Fig. 12A: correlation coefficient $r=0.25$, paired t-test, $p=0.01$). The same is true for the theoretical observer (B: $r=0.44$, paired t-test, $p=0.0001$).

3. Discussion

The main findings in this study can be summarized as follows. First, we demonstrated that spike timing is a source of information — in addition to spike counts — that can potentially contribute to the coding of stimulus contrast in area MT neurons. Second, we validated previous reports of the contribution of spike timing to coding of stimulus direction in MT neurons. Third, we demonstrated that discrimination of motion direction and contrast by a theoretical observer model improves when using metrics that incorporate both, spike counts, and the information provided by spike timing relative to using spike counts only. Fourth, different from previous studies in MT, we applied and validated in our sample of neurons the metric-based approach proposed by Victor and Purpura (1996) to determine the potential role of spike timing in information coding.

3.1. *Contribution of spike timing to information coding in MT neurons*

Both the information and signal detection theoretic approaches used in our study provide complementary insights into neural coding, and they have also both been previously used to quantify the stimulus-encoding capabilities of visual cortical neurons (Bair et al., 1994; Newsome et al., 1989; Reich et al., 2001; Samonds and Bonds, 2004). While the mutual information approach provides a measure of stimulus clustering that applies to a metric space, and is independent of any assumed geometry of responses to stimuli, the signal detection (theoretical observer model) approach treats spike train distances as vectors within a Euclidean space, mimicking the performance of a decision-maker that has access to the spike trains produced by MT neurons.

So far, in area MT neurons, the use of the metric-space approach of Victor and Purpura (1996) has not been reported. We obtained a significant increase in information content in the experimental conditions tested (in both animals) using either the D^{spike} or the $D^{product}$ metric compared to using D^{count}. Furthermore, the theoretical observer analysis suggests that this increase in information occurred more reliably using D^{spike}, translating into almost 10% improvement in discrimination performance (in the range from 55% to 82%). Although this improvement may appear modest, one must consider that a 10% increase in discrimination performance within the mentioned range could have a profound impact on behavior. It is also possible that our theoretical observer analysis underestimated the contribution of the increase in information content provided by spike timing to behavior. Nevertheless, we have convincingly demonstrated improved performance when an observer takes into account spike timing in addition to spike count information.

3.2. *Interpretation of the metrics behavior*

Unlike D^{count}, both D^{spike} and $D^{product}$ measure similarity of temporal structure between spike trains. Overall, however, the $D^{product}$ metric provided smaller improvements in performance compared to D^{count} than D^{spike}. To better understand this result, we repeated our analysis on two types of spike trains generated by the PMPD and spike jitter models, with a known temporal structure. In the PMPD model, temporal frequency or temporal phase is used as a coding parameter. In the spike jitter model, random (Poisson) spike trains with no predominant oscillatory structure are identical within categories except for a uniformly distributed temporal jitter.

The potential for superiority of $D^{product}$ at detecting differences in frequency and phase is intuitive since it is a vector space computation and the deterministic components of the PMPD model for different values of f_0 are orthogonal. Similarly, a metric such as D^{spike}, in which cost is associated with spike shifting, ought to perform best for spike trains differing by random jitter of spike times. This is demonstrated in Fig. 11E. D^{spike} appears to perform better than $D^{product}$ when spike timing information, not based on changes in periodicity, exists. Conversely, $D^{product}$ tended to perform better than D^{spike} when spike timing information is based on changes in periodicity. Note that the metric D^{spike} also contains information about periodicity in a spike train (as demonstrated in Figs. 11A–D), however, in that respect $D^{product}$ is superior.

The overall better performance of D^{spike} over $D^{product}$ in the theoretical observer analysis of the recorded spike trains suggests that the information embedded in the temporal structure of area MT neuronal spike trains (at least in our experiments) is not due to periodic oscillations. This interpretation is consistent with the findings of Bair et al. (1994). Additionally, D^{spike} contains information about spike counts, while $D^{product}$ is insensitive to this parameter. Thus, the superiority of D^{spike} is due to the fact that it captures both, counts and timing information.

3.3. *Comparison with previous studies in area MT*

Our results agree with previous reports of the spike timing contribution to coding of motion attributes (speed and direction) by MT neurons (Bair and Koch, 1996; Buracas et al., 1998; Fellous et al., 2001; Masse and Cook, 2008; Osborne et al., 2004). Importantly, we expand this previous work and demonstrate that spike timing also contributes to coding of stimulus contrast.

In a previous study, Bair and Koch (1996) computed spike jitter standard deviation between spike trains evoked by identical stimuli (RDPs with varying degrees of coherence for dots moving in the same direction of motion). The highest temporal precision (lowest spike jitter standard deviation) was evoked by low coherence RDPs, and was practically absent for 100% coherence patterns. In our study, we used 100% coherence RDPs, and found increased information content when using D^{spike}, a metric that incorporates spike timing and count information. There are at least two differences between our methodology and the one of Bair and Koch (1996). First, we used a metric-based analysis while they used spike jitter standard deviation measurements. The metric approach may be more sensitive to subtleties in temporal structure of the spike trains. Second, Bair and Koch (1996) analyzed until 2 sec of stimulus presentation — at such a long interval from stimulus onset we found a negligible effect of spike time. In addition to the latter point, the study reporting contribution of spike timing to contrast coding in V1 neurons analyzed (as we did) response periods shortly following stimulus onset (Reich et al., 2001). Thus, the choice of a response period seems to have an impact when quantifying the potential contribution of spike timing to coding.

In another study, Fellous et al. (2004) reported evidence for temporal structure in the firing pattern of MT neurons. They used the normalized inner product of the spike density function (Schreiber, 2003) as their measure of similarity. However, they used a vector based clustering algorithm, and therefore required the additional step of histogram-reshaping with a sigmoid function. The fact that they found substantially improved performance using this measure, whereas the improvement we found seem to be smaller, may be a consequence of the stimulus used. In that study, the animal viewed moving Gabor patches that may have elicited a higher oscillatory response by stimulus-locking due to the spatio-temporal frequency component of the patch. If that were the case, as our PMPD simulations demonstrated, we may expect a high performance from a metric based on an inner product such as $D^{product}$. Thus, the choice of a stimulus seems to have an impact in the type of timing information (periodic vs. non-periodic) that contributes to coding.

Osborne et al. (2004) also analyzed the information content of direction selective responses of MT neurons (using firing rate) as a function of time. They found that the first 100 msec of the spike train after the response latency (~80 msec) contains 80% of the information about the stimulus direction. Our results agree with this report and further suggest that even more stimulus information can be gleaned within this time interval by making use of temporal precision. Interestingly, there is psychophysical evidence that temporal integration of motion direction by humans, in tasks using highly coherent RDPs, does not dramatically change after 250 msec of stimulus presentation (Watamaniuk and Sekuler, 1992). Considered together, the results of these two studies may suggest that spike timing information is mainly available

during the time following the stimulus appearance, in which a perceptual judgment about the stimulus is likely made.

Other studies in area MT of macaques have found changes in the synchronized activity of neurons shortly before the animals make decisions (Thiele and Hoffmann, 2008). Although our study did not directly examine this type of temporal relationship between the firing of MT units, the superiority of a metric such as D^{spike} that contains timing and rate information may be related to these findings. Timing information that is not strictly associated to a given frequency may underlie changes in synchrony between the firing of populations of MT neurons over short time scales. This issue, however, needs to be further investigated using simultaneous recordings of neuronal activity in the behaving animal.

3.4. *Theoretical observer model analysis*

In order to demonstrate the practical advantage of temporal coding during perceptual tasks, we introduced a theoretical observer model. The observer plays the role of a decision-maker that has access to the output of the recorded MT neuron and compares it to a template in order to solve the discrimination task. One potential problem with this method is choosing the template. In our analysis, we estimated the distance (e.g. D^{spike}) from each recorded spike train to the theoretical template using only recoded spike trains (see Appendix A and Experimental procedure — Mutual information of stimulus clustering). It is difficult with our data to speculate on potential physiological correlates for a template representation of a time-varying signal. Some possibilities are: the coincidence detection of spatially correlated spikes, variations in the weighing of synaptic decays within a dendritic field, and/or different configurations of time-varying membrane voltage channels (Di Maio, 2008; Mockett and Hulme, 2008). These ideas, however, must be experimentally tested.

3.5. *Temporal features of spike trains*

We investigated whether some of the temporal features of the spike trains, such as response latency, or periodicity in the timing of spikes, could fully explain the advantage seen with D^{spike} relative to the other metrics. The data shown in Figs. 10 and 11 demonstrate that this alone is not the case. One likely explanation for our results is that more than one feature of spike trains — such as a hybrid of spike count, latency, and possibly the shape of the transient "envelope" (to which $D^{product}$ would be sensitive) may have produced the increase in information and performance seen with the D^{spike} metric. Thus, any method that would examine one of these individual features in isolation will necessarily underperform D^{spike}, and likely underestimate how the brain uses different types of available information.

In conclusion, this study used a metric-based analysis to demonstrate the potential advantage in the use of the spike timing information embedded in spike trains fired by area MT neurons when discriminating stimuli that vary in contrast and direction. Although our study cannot demonstrate that the brain is indeed using spike times to boost discrimination performance, our results would explain some differences between neurometric and subjects' performance found in studies using spike counts to quantify information content in spike trains fired by visual neurons (Cohen and Newsome, 2008; Newsome et al., 1989; Rodriguez-Sanchez et al., 2009; Snowden and Braddick, 1990; Snowden et al., 1991).

4. Experimental procedures

4.1. *Stimuli*

The data reported in Experiment 1 are part of a larger data set from another study (Khayat et al., 2010a). Stimuli consisted of moving RDPs presented on a rear projection screen using a video projector (NEC WT610, NEC Inc., Japan). The screen resolution was 1024×768 with a refresh rate of 85 Hz. Each pixel was 0.083 cm^2 and each dot size was 0.17 cm^2. The monkey's eyes were 57 cm away from the monitor. At this distance, each pixel approximated 0.08° of visual angle.

Within a pattern, the dots moved behind a circular aperture with 100% coherence at the preferred speed of the recorded neuron, and had infinite lifetime. In different trials, the initial dots' position (seed) varied, although the RDP's dot density remained constant. This avoided using the same seed within a stimulus type, and different seeds across stimuli, which may produce specific patterns of spike timing corresponding to each stimulus type. When a dot crossed the aperture's border, it was replotted on the opposite side avoiding dots' appearance and disappearance at random positions within the aperture, which may cause variable amounts of flicker depending on the dot's speed.

Importantly, unlike previous studies in area MT (e.g., Masse and Cook, 2008) our RDPs had 100% coherence. They therefore did not contain random fluctuations in the dot statistics due to changes in dots' coherence that may cause changes in the firing pattern of MT neurons. Such fluctuations are known to produce changes in human behavioral thresholds (Barlow and Tripathy, 1997). Rather, our stimuli produced a stable percept of a rigid surface composed of approximately equaled spaced dots moving behind a circular aperture (Martinez-Trujillo and Treue, 2002; Treue and Martinez-Trujillo, 1999).

4.2. *Task*

4.2.1. *Experiment 1*

During trials, the animals fixated (maintained gaze) on a dot (0.3° square) positioned approximately at the screen center. A virtual circular window (invisible to the animal) with 1.5° diameter and centered on the dot defined the area in which if the animal kept fixation for more than 200 msec, it could initiate the stimulus presentation by pressing a button. After 470 msec, two pairs of RDPs appeared on the screen, one pair located inside the RF of the recorded neuron, and the other located outside at a symmetric position in the opposite hemifield (Fig. 13A). After a random delay (1480–3670 msec), the fixation dot changed luminance and the monkey had to release the button within a 350 msec response time window,

starting 150 msec after the change, in order to obtain a reward (drop of water or juice). Trials in which the monkey broke fixation before releasing the button, or released the button before the fixation dot changed luminance were aborted without reward and considered errors. The animals correctly performed the task in more than 90% of the trials.

Each pair of stimuli consisted of one high contrast RDP [luminance standard deviation = 13 cd/m^2 (Moulden et al., 1990)] moving in the neuron's antipreferred (null) direction, and one test RDP (Fig. 13A). The latter could have two different configurations across trials: a) it could move in the neuron's preferred direction with trial-to-trial changes in contrast (100%, 14%, 1.5%, 0.7%, 0.3%, 0.1%, and 0.02% relative to the high contrast stimulus), or b) it could have the same contrast as the null pattern but with trial-to-trial variations in its motion direction (in intervals of 15° departing from the preferred direction until 90° away from that direction). The contrast and direction values were chosen in such a manner that they produced similar response variations along the neurons' contrast and direction response functions [see Fig. 1, (Martinez-Trujillo and Treue, 2002; Treue and Martinez-Trujillo, 1999; Khayat et al., 2010a)].

4.2.2. *Experiment 2*

In order to control for any potential impact that the stimulus configuration might have had on temporal coding, we analyzed a second data set from another experiment (Niebergall et al., 2008). The task and timing of the trials was similar to Experiment 1, however, there was only one RDP inside the neurons RF. Additionally, the direction of motion of the RDP changed between trials in steps of 30° (rather than in steps of 15% as in Experiment 1), departing from the preferred direction (Fig. 13B). The two animals participating in this experiment were the same as in Experiment 1; however, because the differences in the stimulus configuration and the fact that Experiment 2 was conducted approximately one year later we decided to analyze the data from both experiments separately.

4.3. *Animal preparation*

Two *Macaca mulatta* (Se and Lu) were used in the experiments. Each animal was surgically prepared for single unit recordings as previously described (Martinez-Trujillo and Treue, 2004). Briefly, during each experimental session, the head was fixed with a surgically implanted head post, and eye movements were tracked using a video based eye tracker (Eyelink II, SR Research Ltd., Canada). During each experimental session, one or two penetrations were made using tungsten microelectrodes (FHC Inc., USA, impedance = 0.5–3 MΩ), and an electric microdrive (Plexon Inc., USA) positioned on the top of a recording chamber implanted over the right parietal bone. Structural magnetic resonance images were used to localize the area of recording (Fig. 13C), and neurons were classified as MT units based on their response properties (receptive field size and tuning for linear motion direction and spiral stimuli) (Martinez-Trujillo and Treue, 2004). Single unit responses were isolated and sorted online and offline using Plexon spike sorting software (Plexon Inc., USA), which allowed extracting the times of occurrence of each spike. Fig. 1 shows spike trains and PSTHs of a typical MT unit to stimuli with varying contrast and direction. These protocols and procedures were in accordance with Canadian Council on Animal Care guidelines and were pre-approved by the McGill University Animal Care Committee.

4.4. *Spike train metrics*

We analyzed spike trains using the metric-space method of Victor and Purpura (1996, 1997). Three metrics were considered: D^{count}, $D^{spike}(q)$, and one we refer to as $D^{product}(\sigma)$. These metrics produce measurements of the distance between two spike trains in terms of the "cost" of transforming one spike train into the other. In D^{count}, the simplest of the three metrics, there is a cost of 1 for every spike added or removed, but no cost for moving a spike in time. This reduces to the arithmetic difference in firing rates ($r_2 - r_1$), and thus implies a simple rate code in which the timing of spikes is irrelevant.

In $D^{spike}(q)$, there is a cost of 1 for adding or removing a spike (as in D^{count}), but there is also a cost, q, associated with moving a spike in time. The total "distance" from one spike train to the other is the sum of the cost associated with one of two elementary steps required to transform one spike train into the other: $q \cdot \Delta t$, the cost of shifting a spike in time, or 1, the cost of adding or deleting each spike. The parameter q is in sec^{-1}. The use of these two elementary steps is governed by the following rule: if two spikes in different spike trains are separated in time by an interval greater than $2/q$ (in seconds), then it is less costly to simply delete and replace the spike at the second location, and this operation is done at a cost of 2. Thus, q reflects temporal precision since it can be thought of as the difference in the timing of occurrence of two spikes that makes just as much difference to the nervous system as a spike deletion (Di Lorenzo and Victor, 2003). $D^{spike}(q \to 0) = D^{count}$. Therefore, the first point on the $D^{spike}(q)$ information function reduces to D^{count}. We used the spike toolkit (Goldberg et al., 2005) to calculate $D^{spike}(q)$ and the mutual information.

The third spike train metric is based on the vector product between spike trains (Schreiber, 2003), and in keeping with the Victor and Purpura (1997) notation, we refer to it as $D^{product}$. Each spike train is convolved with a Gaussian filter, with size (kernel) σ, to form a spike density function s_i. We define the distance between two spike trains as:

$$D^{product}(\sigma) = 1 - \frac{\langle s_i, s_j \rangle}{||s_i|| \cdot ||s_j||},$$

where σ is the Gaussian filter size parameter. The intuition behind this metric is that spike trains are considered similar if the normalized product of their spike density functions is high.

4.5. *Theoretical observer model*

In this section, we illustrate the strategy of a theoretical observer model subject to a metric-based strategy. The task for the model is to determine whether a spike train was elicited by one of two different stimuli. For simplicity, the analysis is restricted to a 2-alternative forced choice (2AFC) discrimination task. Consider two equiprobable signals, $S(\theta_0)$ and $S(\theta_1)$, where θ is the modulated stimulus parameter. One

of these signals, $f(\theta_0,\theta_1)$, enters the visual system and "arrives" to area MT, where a neuron produces a spike train, $r(f(\theta_0,\theta_1))$. The theoretical observer "reads" the signal and assigns it to one of two categories corresponding to the different stimuli.

The observer compares the generated spike train to two templates, $\bar{r}(S(\theta_0))$ or $\bar{r}(S(\theta_1))$, each representing the "typical" spike train generated by each of the two stimuli. Metric-based distances, $D[\bar{r}(S(\theta_0)), r(f(\theta_0,\theta_1))]$ and $D[\bar{r}(S(\theta_1)), r(f(\theta_0,\theta_1))]$ are calculated using D^{count}, D^{spike}, or $D^{product}$. For each metric, the stimulus–template combination generating the minimum distance determines the most likely stimulus that generated the spike train. This method is somewhat similar to the signal detection approach used by previous studies to quantify neurometric performance in MT neurons (Bair et al., 1994; Newsome et al., 1989). However, instead of using some measure of neuronal output strength, it uses spike train metrics.

One potential difficulty with this method is that the spike train templates $\bar{r}(S(\theta_0))$ and $\bar{r}(S(\theta_1))$ are unknown. Thus, we obtained an estimate of such templates from distributions of spike train distances within categories and between categories (see Appendix A for details). This system is ideal in the sense that it assigns a spike train to the most likely group of spike trains from which it belongs, according to the different metrics. The method is thus useful for comparing spike train metrics to one another using real data. The formal derivation of the theoretical observer's performance, $P_{correct}$, appears in the Appendix A.

4.6. *Metric-based data analysis*

The metrics are applied to the data set in the following way. First, the spike trains are numbered such that if there are 2 stimuli, and 8 spike trains recorded in response to each, then the spike trains would be ordered from 1 to 16. For each metric parameter, a square matrix of distances between each spike train (i.e. 16×16 elements in this example) is computed, where the value of each element represents the distance between 2 spike trains according to the metric used at the value of the parameter tested (Fig. 2, step 2). This step is repeated for different values of the metric's free parameter, q or σ. A stimulus-dependent clustering method (Victor and Purpura, 1997) is then used to calculate the mutual information, H, in bits.

This method does not assume that spike train distances lie in a vector space, and makes no assumptions about parametric relationships between stimulus classes. The mutual information, H, is then plotted against the free parameter of the metric being used (Fig. 2, step 3). To correct for bias associated with the direct calculation of information (Treves and Panzeri, 1995), we used the re-sampling technique in which 10 recalculations of the transmitted information, H, are done following random reassignments of the observed responses to the stimulus classes. The average of these values is used as the estimate of bias due to chance clusters, and it is subtracted from the mutual information estimate of each neuron (Victor and Purpura, 1997).

The value used for D^{count} is the value at $q=0$ on the D^{spike} curve (Fig. 2, step 3, black curve on the left panel). The values used for D^{spike} and $D^{product}$ were the maximum value on the corresponding curves (Fig. 2, step 3, black curve on the right panel). Error bars were generated using the bootstrap method in which stimulus class indices were randomly selected, and mutual information (bias-corrected) and the theoretical observer's performance were recomputed. One hundred such recalculations were done for each point, and confidence intervals were computed using the bias-corrected and accelerated percentile method (Efron and Tibshirani, 1993). The bootstrap estimates were also used for significance testing for each individual neuron using the Wilcoxon rank-sum test to allow comparisons between the metrics. In Experiment 1, we recorded 5–12 trials per stimulus. We analyzed the initial 200 msec window following stimulus onset in 102 neurons from two animals. In Experiment 2, we recorded 4–11 trials per stimulus, and analyzed the initial 200 msec window following stimulus onset in 103 neurons from the same two animals.

4.7. *Latency analysis*

We determined the latency of activity onset in each spike train using a method motivated by Maunsell and Gibson (1992). First, we obtained a set of estimates of the spontaneous or baseline firing rate using the 300 msec pre-trial interval. Since we desired a unique estimate for each spike train, we could not use the peristimulus time histogram (PSTH) constructed from pooling spike trains. Instead, we represented the spike trains using time bins with duration τ. If the estimated firing rate within a time bin was statistically higher than the baseline (using t-test, $p<0.001$), this was determined to be the onset of activity. For small values of τ, this would reduce to the time to the first spike (since, for example, at 1 msec resolution a single spike would yield a firing rate of 1000 Hz). For very large values of τ, temporal resolution limited the ability of the theoretical observer to make decisions. Estimates of latency for a range of values, τ=5, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100 msec, were obtained.

In order to allow comparison with D^{spike}, D^{count}, and $D^{product}$, we used the differences between latencies to compute $D^{latency}$. Thus for each pair of spike trains (see Fig. 2), the difference between the two latencies is considered to be $D^{latency}$, a confusion matrix is calculated, and the theoretical observer (or alternatively, information) is computed. For each stimulus pair used in the previous analyses (preferred and non-preferred), we selected for each neuron the τ-value that maximized the performance of the theoretical observer.

We also characterized latency and computed $D^{latency}$ performance using a different method, which does not make use of time binning. First, we fit a 3rd order polynomial to the response from time 0 to 200 msec. Then, we used one of four methods to determine latency of activity onset (Fig. 10D). In the first method the first threshold cross is determined to be the latency. The threshold was estimated as the mean firing rate in the 300 ms pre-trial interval plus 1.5 standard deviations (we also used 1, and 2 SD with equally poor results). The second method uses the first positive inflection point of the fit as the latency. The third method uses the maximum value of the fit as the latency. The fourth method was a hybrid of these three methods in which the threshold was used unless it is not reached or it is the initial value. In

that case the maximum is used unless it is the initial or final value (i.e., not a local maximum). In that case, the first inflection point is used. All of these methods faired poorly compared to the binned method that we used (see performance measurements in Figs. 10C and D). One possible explanation is that these methods are better suited for use with post-stimulus time histograms computed from multiple stimulus presentations than for single spike trains.

4.8. *Metric assessment using simulated data*

To better understand the behavior of the two metrics, D^{spike} and $D^{product}$, we applied the analysis to two sets of simulated data with known properties. The first simulation is based on frequency modulation. In this simulation, spike trains are generated from an adaptation of the periodically modulated Poisson distribution (Hunter and Milton, 2003), in which the instantaneous firing rate $\lambda(t)$ is defined as:

$$\lambda(t) = \begin{cases} 0 & \text{if } m\sin(2\pi f_m t) < 0 \\ \lambda_0(1 + m\sin(2\pi f_m t)) & \text{if } 0 \le m\sin(2\pi f_m t) \le 1 \\ 2\lambda_0 & \text{if } m\sin(2\pi f_m t) > 1 \end{cases},$$

where λ_0 is the carrier rate, m is the modulation amplitude, and f_m is the modulation frequency. Note that this function is equivalent to a sine function for $0<m<1$, and approaches a square wave as $m\to\infty$. From this distribution, we generated two sets of spike trains representing responses from two theoretical stimuli. The sets of spike trains differed only in the parameter, f_m. We tested this model using two sets of frequency parameters. The low frequency parameters ($f_1=5$ Hz, $f_2=15$ Hz) were chosen in part because of the observation of a strong peak in the low frequency range in LFP's recorded from areas MT (Khayat et al., 2010b) and LIP (Pesaran et al., 2002), and in part because of the low frequency stable spiking patterns elicited in pyramidal cells *in vitro* (Fellous et al., 2004). We also tested this model at higher frequency parameters ($f_1=40$ Hz, $f_2=50$ Hz) motivated by the 30–70 Hz synchronous oscillations found in adjacent neurons in macaque area MT (Kreiter and Singer, 1992), and by the 40 Hz peak in the power spectrum of spike trains recorded from area MT (Bair et al., 1994). In these simulations, increasing m makes the task of clustering spike trains to the parent distribution easier.

In order to simulate phase encoding within spike trains (Latham and Lengyel, 2008; Montemurro et al., 2008), we repeated the periodically modulated Poisson distribution (PMPD) analysis at low frequency ($f_1=f_2=15$ Hz) and at high frequency ($f_1=f_2=50$ Hz) but with spike trains from each category offset by a phase of $\Pi/4$.

The second simulation is based on spike jitter, with no significant difference in the frequency spectra between the two sets of spike trains. In this simulation, two template spike trains are generated from a Poisson process with the same rate parameter, λ. From each template, a set of spike trains is generated by jittering each spike with a Gaussian offset. Formally, let $\{g_1,\ldots,g_n\}$ represent event times ~ Poisson(λ). In the first set of spike trains, the jth simulation produces a spike train $g(j)=\{X_{j1},\ldots,X_{jn}\}$, where the element $X_{ji}\sim\text{norm}(g_i,\sigma)$. Thus the spike train $g(j)$ is equivalent to g (a Poisson spike train), except for a Gaussian jitter applied to each spike. Similarly let $\{h_1,\ldots,h_m\}$ represent event times ~ Poisson(λ). In the second set of spike trains, the jth simulation produces a spike train $h(j)=\{Y_{j1},\ldots,Y_{jm}\}$, where the element $Y_{ji}\sim\text{norm}(h_i,\sigma)$. σ is the free parameter of the model representing the amount of spike jitter in seconds. For very small values of σ, classification is trivial because all of the spike trains belonging to the $g(j)$ group would be very similar to each other, and all of the spike trains belonging to the $h(j)$ group would be similar to one another for the same reason. Increasing σ makes the task of clustering spike trains to the parent distribution, $\{g_1,\ldots,g_n\}$ or $\{h_1,\ldots,h_m\}$, more difficult.

In all cases, 100 simulations were done, each using fifteen trials of 200 msec, with the baseline firing rate, $\lambda_0=40$ Hz. We then applied the three metrics, D^{spike}, $D^{product}$, and D^{count} to these data sets and compared their performance as defined in the previous section. The simulations were conducted over a range of parameters: m in the PMPD simulations and σ for the spike jitter simulations.

4.9. *Mutual information of stimulus clustering*

We used the method of Victor and Purpura (1996, 1997) to quantify the mutual information using spike train metrics. This method is summarized here. Suppose that N_{tot} spike trains were elicited from one of the stimulus classes $s_1,s_2,\ldots,s_c$. Each response can be classified into one of C response classes $r_1, r_2,\ldots,r_c$. This classification can be represented by the matrix $N(s_\alpha, r_\beta)$, whose entries denote the number of times a stimulus s_α elicits a response in class r_β.

Initially set all of the elements of the matrix $N(s_\alpha, r_\beta)$ to zero. For each stimulus class s_γ, calculate $d(S,s_\gamma)$, the average distance from S to the spike trains elicited in stimulus class s_γ as:

$$d(S,s_\gamma) = \left[\left\langle (D[q](S,S'))^z \right\rangle_{S'}\right]^{1/z},$$

where the operator $\langle\rangle_{S'}$ denotes the mean over all values of S', and S' is the set of all spike trains elicited by the stimulus class s_γ. Note that spike train self-comparisons are excluded from this calculation (i.e. $D[q](S,S)$). The spike train S is then classified into the response class r_β that minimizes the average distance $d(S, s_\gamma)$, and $N(s_\alpha, r_\beta)$ is incremented by 1. If the minimum is shared by k-distances then each element is incremented by $1/k$. In the information analyses presented here, $z=-2$. However, changing the value of z to 1 did not change the results. For example, in the contrast task of Experiment 1, the mean information (bias-corrected, both monkeys pooled, using $z=1$) was 0.28 using D^{count} and 0.65 using D^{spike}, and in the direction task was 0.27 D^{count} and 0.61 for D^{spike}. In the theoretical observer analysis, $z=1$ was used (see Appendix A).

The information, H, is used to quantify the extent to which $N(s_\alpha, r_\beta)$ is not random: $H = \frac{1}{N_{tot}}\sum_{\alpha,\beta} N(s_\alpha, r_\beta)\left[\log_2 N(s_\alpha, r_\beta) - \log_2\sum_a N(s_\alpha, r_\beta) - \log_2\sum_b N(s_\alpha, r_\beta) + \log_2 N_{tot}\right]$.

The formula above may be contrasted against the probability of a correct response of the theoretical observer, derived in the Appendix A but represented here in the Victor and Purpura notation for simplicity:

$$P_{correct} = 0.5\cdot\frac{N(s_1,r_1)+N(s_1,r_2)}{N(s_1,r_1)} + 0.5\cdot\frac{N(s_2,r_2)+N(s_2,r_1)}{N(s_2,r_2)}.$$

Acknowledgments

We thank Walter Kucharski for his technical assistance. We thank Dr. Eric Cook, and Dr. Nick Masse for their useful comments on the manuscript.

Grants: This research was supported by grants from the CFI, CRC, CIHR, and The Ottawa Hospital Department of Surgery, and The Ottawa Hospital Division of Neurosurgery.

Appendix A

The performance of a theoretical observer subject to a metric-based strategy is calculated here. This observer is ideal in the sense that it makes the best possible decision subject to a metric-based strategy. It is theoretical in that its derivation assumes the existence of a spike train template. First, let us consider a general mathematical form of the stimulus,

$$f_i(\theta_0,\theta_1) = G_i\cdot\theta_1 + (1-G_i)\cdot\theta_0,$$

where G_i is a random variable with binomial distribution assuming a value of 0 or 1 with equal probability, and i is the trial number. The neuron's response can be described by the response function $r_i(f(\theta_0,\theta_1))g_i(f(\theta_0,\theta_1))$. In this model, the system uses template spike trains, $\bar{g}(S(\theta_0))$, and $\bar{g}(S(\theta_1))$. We define the decision that the system makes in terms of a decision function, where D^{metric} is the metric being tested, and may represent D^{count}, D^{spike}, $D^{product}$ or any other spike

$$dec_{D^{metric}}(f(\theta_0,\theta_1)) = \begin{cases} 0, & \text{if } P[G=0|f(\theta_0,\theta_1)] > P[G=1|f(\theta_0,\theta_1)]; \\ 1, & \text{if } P[G=1|f(\theta_0,\theta_1)] > P[G=0|f(\theta_0,\theta_1)]; \end{cases}$$

train metric. To derive the probability of this decision-maker being correct, $P_{correct}$, we first derive the probability of an incorrect response, $P_{incorrect}$:

$$\begin{aligned} P_{incorrect} &= P(dec_{D^{metric}}(f(\theta_0,\theta_1))\neq G) \\ &= P(G=1)\cdot P[dec_{D^{metric}}(f(\theta_0,\theta_1)) = 0|G=1] \\ &\quad + P(G=0)\cdot P[dec_{D^{metric}}(f(\theta_0,\theta_1)) = 1|G=0] \\ &= P(G=1)\cdot P[D^{metric}\{g_i(f(\theta_0,\theta_1)),\bar{g}(S(\theta_1))\} \\ &\quad > D^{metric}\{g_i(f(\theta_0,\theta_1)),\bar{g}(S(\theta_0))\}|G=1] \\ &\quad + P(G=0)\cdot P[D^{metric}\{g_i(f(\theta_0,\theta_1)),\bar{g}(S(\theta_1))\} \\ &\quad \le D^{metric}\{g_i(f(\theta_0,\theta_1)),\bar{g}(S(\theta_0))\}|G=0] \end{aligned}$$

To simplify the derivation, let us introduce the following notation. Let $d_{r_i,0} = D^{metric}\{g_i(f(\theta_0,\theta_1)),\ \bar{g}(S(\theta_0))\}$, that is, the distance using a given metric between the real spike train and the template spike train for the stimulus $S(\theta_0)$, for each trial, i. Similarly, let $d_{r_i,1} = d^{metric}\{g_i(f(\theta_0,\theta_1)),\ \bar{g}(\theta_1)\}$. We then define the set of all such distances with the first template, $D_{r,0} = \{d_{r_1,0}, d_{r_2,0},\ldots,d_{r_n,0}\}$ and with the second template, $D_{r,1} = \{d_{r_1,0}, d_{r_2,0},\ldots,d_{r_n,0}\}$. Also assume that the stimuli are equiprobable, that is $P(G=1)=P(G=0)=0.5$. Then for the ith trial,

$$\begin{aligned} P_{i,incorrect} &= 0.5\cdot\left\{P\left[d_{r_i,0} < d_{r_i,1}|G=1\right] + P\left[d_{r_i,0} > d_{r_i,1}|G=0\right]\right\} \\ &= 0.5\cdot\left\{P\left[d_{r_i,0}-d_{r_i,1} < 0|G=1\right] + P\left[d_{r_i,0}-d_{r_i,1} > 0|G=0\right]\right\}. \end{aligned}$$

One may use all of the stimulus responses to estimate $P_{incorrect}$ as:

$$P_{incorrect} = 0.5\cdot\left[\frac{1}{m}\sum_{i=1}^{m}\begin{cases}1, & \text{if } d_{r_i,0}-d_{r_i,1}>0\\ 0, & \text{if } d_{r_i,0}-d_{r_i,1}\le 0\end{cases} + \frac{1}{n}\sum_{i=m+1}^{m+n}\begin{cases}1, & \text{if } d_{r_i,0}-d_{r_i,1}<0\\ 0, & \text{if } d_{r_i,0}-d_{r_i,1}\ge 0\end{cases}\right],$$

where the first m spike trains are in response to stimulus $S(\theta_0)$, and spike trains $m+1$ to n are in response to stimulus $S(\theta_1)$.

The distances in the above expression, $d_{r_i,0}$ and $d_{r_i,1}$, cannot be computed directly since the templates, $\bar{r}(\theta_0)$ and $\bar{r}(\theta_1)$, are unknown. We estimate the above expression by using each spike train as a template and calculating cross metrics between spike trains. To do this we assume that the template is a centroid within the set of measured responses and define the set of distances:

$$\begin{aligned} \widehat{d}_{11} = \{&D^{metric}[r_1,r_2], D^{metric}[r_1,r_3],\ldots, D^{metric}[r_1,r_m], D^{metric}[r_2,r_3],\ldots, \\ &D^{metric}[r_2,r_m],\ldots, D^{metric}[r_{m-1},r_m]\}, \end{aligned}$$

where m is the number of trials with stimulus $S(\theta_0)$. Note that this set has $\alpha = \sum_{i=2}^{m}\sum_{j=1}^{i-1} 1$ elements.

Similarly,

$$\begin{aligned} \widehat{d}_{10} = \{&D^{metric}[r_{m+1},r_1], D^{metric}[r_{m+1},r_2],\ldots, D^{metric}[r_{m+1},r_m],\ldots, \\ &D^{metric}[r_{m+n},r_1],\ldots, D^{metric}[r_{m+n},r_m]\}, \end{aligned}$$

where n is the number of trials with stimulus $S(\theta_1)$. This set has $\beta = mn$ elements. Finally,

$$\begin{aligned} \widehat{d}_{00} = \{&D^{metric}[r_{m+2},r_{m+1}], D^{metric}[r_{m+3},r_{m+1}],\ldots, \\ &D^{metric}[r_{m+3},r_{m+2}],\ldots, D^{metric}[r_{m+4},r_{m+1}],\ldots, \\ &D^{metric}[r_{m+4},r_{m+3}],\ldots, D^{metric}[r_{m+n},r_{m+1}],\ldots, \\ &D^{metric}[r_{m+n},r_{m+n-1}]\}, \end{aligned}$$

which contains $\gamma = \sum_{i=m+2}^{m+n}\sum_{j=m+1}^{i-1} 1$ elements. The estimate of $P_{incorrect}$ based on the recorded spike trains is:

$$\hat{P}_{incorrect} = \frac{1}{2}\left[\frac{1}{\beta\gamma}\sum_{i=1}^{\beta}\sum_{j=1}^{\gamma}\begin{cases}1, & \text{if } \hat{d}_{10}(i)-\hat{d}_{11}(j)<0\\ 0, & \text{if } \hat{d}_{10}(i)-\hat{d}_{11}(j)\ge 0\end{cases} + \frac{1}{\alpha\beta}\sum_{i=1}^{\alpha}\sum_{j=1}^{\beta}\begin{cases}1, & \text{if } \hat{d}_{10}(i)-\hat{d}_{00}(j)<0\\ 0, & \text{if } \hat{d}_{10}(i)-\hat{d}_{00}(j)\ge 0\end{cases}\right].$$

Finally, the probability of a correct response of the theoretical observer subject to a metric-based analysis is $\hat{P}_{correct} = 1-\hat{P}_{incorrect}$. We refer to this entity throughout the text as $P_{correct}$.

REFERENCES

Bair, W., Koch, C., 1996. Temporal precision of spike trains in extrastriate cortex of the behaving macaque monkey. Neural Comput. 8, 1185–1202.

Bair, W., Koch, C., Newsome, W.T., Britten, K.H., 1994. Power spectrum analysis of bursting cells in area MT in the behaving monkey. J. Neurosci. 14, 2870–2892.

Barlow, H., Tripathy, S.P., 1997. Correspondence noise and signal pooling in the detection of coherent visual motion. J. Neurosci. 17, 7954–7966.

Buracas, G.T., Zador, A.M., DeWeese, M.R., Albright, T.D., 1998. Efficient discrimination of temporal patterns by motion-sensitive neurons in primate visual cortex. Neuron 20, 959–969.

Cohen, M.R., Newsome, W.T., 2008. Context-dependent changes in functional circuitry in visual area MT. Neuron 60, 162–173.

Di Lorenzo, P.M., Victor, J.D., 2003. Taste response variability and temporal coding in the nucleus of the solitary tract of the rat. J. Neurophysiol. 90, 1418–1431.

Di Maio, V., 2008. Regulation of information passing by synaptic transmission: a short review. Brain Res. 1225, 26–38.

Dubner, R., Zeki, S.M., 1971. Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey. Brain Res. 35, 528–532.

Efron, B., Tibshirani, R., 1993. An Introduction to the Bootstrap. Chapman and Hall/CRC, Boca Raton.

Fellous, J.M., Houweling, A.R., Modi, R.H., Rao, R.P., Tiesinga, P.H., Sejnowski, T.J., 2001. Frequency dependence of spike timing reliability in cortical pyramidal cells and interneurons. J. Neurophysiol. 85, 1782–1787.

Fellous, J.M., Tiesinga, P.H., Thomas, P.J., Sejnowski, T.J., 2004. Discovering spike patterns in neuronal responses. J. Neurosci. 24, 2989–3001.

Goldberg, D.H., Victor, J.D., Gardner, D., 2005. The spike train analysis toolkit: a component of a computational neuroinformatic resource. Society for Neuroscience, vol 35: 984.19, Washington, DC.

Green, D.M., Luce, M.W., 1975. Parallel psychometric functions from a set of independent detectors. Psychol. Rev. 82, 483–486.

Holt, G.R., Softky, W.R., Koch, C., Douglas, R.J., 1996. Comparison of discharge variability in vitro and in vivo in cat visual cortex neurons. J. Neurophysiol. 75, 1806–1814.

Hu, D., Chelvanayagam, D.K., Vidyasagar, T.R., 2002. Irregularity in cortical firing in vivo cannot be explained by inhibition. Proc. Aust. Neurosci. Soc. 13.

Huang, X., Lisberger, S.G., 2009. Noise correlations in cortical area MT and their potential impact on trial-by-trial variation in the direction and speed of smooth-pursuit eye movements. J. Neurophysiol. 101, 3012–3030.

Hunter, J.D., Milton, J.G., 2003. Amplitude and frequency dependence of spike timing: implications for dynamic regulation. J. Neurophysiol. 90, 387–394.

Khayat, P.S., Niebergall, R., Martinez-Trujillo, J.C., 2010a. Attention differentially modulates similar neuronal responses evoked by varying contrast and direction stimuli in area MT. J. Neurosci. 30 (6), 2188–2197.

Khayat, P.S., Niebergall, R., Martinez-Trujillo, J.C., 2010b. Frequency-dependent attentional modulation of local field potential signals in macaque area MT. J. Neurosci. 30 (6), 2188–2197.

Kreiter, A.K., Singer, W., 1992. Oscillatory neuronal responses in the visual cortex of the awake macaque monkey. Eur. J. Neurosci. 4, 369–375.

Kruse, W., Eckhorn, R., 1996. Inhibition of sustained gamma oscillations (35–80 Hz) by fast transient responses in cat visual cortex. Proc. Natl. Acad. Sci. USA 93, 6112–6117.

Latham, P.E., Lengyel, M., 2008. Phase coding: spikes get a boost from local fields. Curr. Biol. 18, R349–R351.

Liu, J., Newsome, W.T., 2006. Local field potential in cortical area MT: stimulus tuning and behavioral correlations. J. Neurosci. 26, 7779–7790.

Malach, R., Schirman, T.D., Harel, M., Tootell, R.B., Malonek, D., 1997. Organization of intrinsic connections in owl monkey area MT. Cereb. Cortex 7, 386–393.

Martinez-Trujillo, J., Treue, S., 2002. Attentional modulation strength in cortical area MT depends on stimulus contrast. Neuron 35, 365–370.

Martinez-Trujillo, J.C., Treue, S., 2004. Feature-based attention increases the selectivity of population responses in primate visual cortex. Curr. Biol. 14, 744–751.

Masse, N.Y., Cook, E.P., 2008. The effect of middle temporal spike phase on sensory encoding and correlates with behavior during a motion-detection task. J. Neurosci. 28, 1343–1355.

Maunsell, J.H., Gibson, J.R., 1992. Visual response latencies in striate cortex of the macaque monkey. J. Neurophysiol. 68, 1332–1344.

Mockett, B.G., Hulme, S.R., 2008. Metaplasticity: new insights through electrophysiological investigations. J. Integr. Neurosci. 7, 315–336.

Montemurro, M.A., Rasch, M.J., Murayama, Y., Logothetis, N.K., Panzeri, S., 2008. Phase-of-firing coding of natural visual stimuli in primary visual cortex. Curr. Biol. 18, 375–380.

Moulden, B., Kingdom, F., Gatley, L.F., 1990. The standard deviation of luminance as a metric for contrast in random-dot images. Perception 19, 79–101.

Newsome, W.T., Britten, K.H., Movshon, J.A., 1989. Neuronal correlates of a perceptual decision. Nature 341, 52–54.

Niebergall, R., Khayat, P., Martinez-Trujillo, J., 2008. Neural correlates of multiple spotlights of attention in area MT during multiple-object tracking. Society for Neuroscience, vol 317..3, Washington, DC.

Osborne, L.C., Bialek, W., Lisberger, S.G., 2004. Time course of information about motion direction in visual area MT of macaque monkeys. J. Neurosci. 24, 3210–3222.

Pesaran, B., Pezaris, J., Sahani, M., Mitra, P., Andersen, R., 2002. Temporal structure in neuronal activity during working memory in macaque parietal cortex. Nat. Neurosci. 5, 805–811.

Raiguel, S.E., Xiao, D.K., Marcar, V.L., Orban, G.A., 1999. Response latency of macaque area MT/V5 neurons and its relationship to stimulus parameters. J. Neurophysiol. 82, 1944–1956.

Reich, D.S., Mechler, F., Victor, J.D., 2001. Temporal coding of contrast in primary visual cortex: when, what, and why. J. Neurophysiol. 85, 1039–1050.

Rodieck, R.W., Kiang, N.Y., Gerstein, G.L., 1962. Some quantitative methods for the study of spontaneous activity of single neurons. Biophys. J. 2, 351–368.

Rodriguez-Sanchez, A.J., Tsotsos, J.K., Treue, S., Martinez-Trujillo, J.C., 2009. Comparing neuronal and behavioral thresholds for spiral motion discrimination. NeuroReport 20, 1619–1624.

Salinas, E., Sejnowski, T.J., 2001. Correlated neuronal activity and the flow of neural information. Nat. Rev. Neurosci. 2, 539–550.

Samonds, J.M., Bonds, A.B., 2004. From another angle: differences in cortical coding between fine and coarse discrimination of orientation. J. Neurophysiol. 91, 1193–1202.

Schreiber, S., 2003. A new correlation-based measure of spike timing reliability. Neurocomputing 7.

Sclar, G., Maunsell, J.H., Lennie, P., 1990. Coding of image contrast in central visual pathways of the macaque monkey. Vis. Res. 30, 1–10.

Shadlen, M.N., Newsome, W.T., 1998. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. J. Neurosci. 18, 3870–3896.
Snowden, R.J., Braddick, O.J., 1990. Differences in the processing of short-range apparent motion at small and large displacements. Vis. Res. 30, 1211–1222.
Snowden, R.J., Treue, S., Erickson, R.G., Andersen, R.A., 1991. The response of area MT and V1 neurons to transparent motion. J. Neurosci. 11, 2768–2785.
Softky, W.R., Koch, C., 1993. The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. J. Neurosci. 13, 334–350.
Thiele, A., Hoffmann, K.P., 2008. Neuronal firing rate, inter-neuron correlation and synchrony in area MT are correlated with directional choices during stimulus and reward expectation. Exp. Brain Res. 188 (4), 559–577.
Treue, S., Martinez-Trujillo, J.C., 1999. Feature-based attention influences motion processing gain in macaque visual cortex. Nature 399, 575–579.
Treves, A., Panzeri, S., 1995. The upward bias in measures of information derived from limited data samples. Neural Comput. 7, 399–407.
Victor, J.D., Purpura, K.P., 1996. Nature and precision of temporal coding in visual cortex: a metric-space analysis. J. Neurophysiol. 76, 1310–1326.
Victor, J.D., Purpura, K.P., 1997. Metric-space analysis of spike trains: theory, algorithms and application. Network 8, 127–164.
Watamaniuk, S.N., Sekuler, R., 1992. Temporal and spatial integration in dynamic random-dot stimuli. Vis. Res. 32, 2341–2347.
Watson, A.B., 1990. Gain, noise, and contrast sensitivity of linear visual neurons. Vis. Neurosci. 4, 147–157.