Perceptual learning of contrast discrimination and its neural correlates in macaque V4 & V1

Xing Chen

B.A. (Neuroscience), University of Southern California, 2008

Thesis submitted to Newcastle University

for the degree of Doctor of Philosophy

August 2013

BRIEF TABLE OF CONTENTS

DETAILED TABLE OF CONTENTS

Abstract

We make frequent evaluations of subtle contrast differences in our visual environment, and often under challenging illumination conditions, whether photopic, scotopic or mesopic. Our contrast discrimination abilities are rigorously honed from an early age, and we continue to carry out these fine perceptual judgments throughout our lifetimes. Thus, the issue of whether substantial improvement in contrast discrimination is possible during later periods in life, such as during adulthood- and the circumstances that allow this- has sometimes come under discussion.

Our adult macaque subjects underwent extensive training on a contrast discrimination task, in which stimuli were positioned at a variety of peripheral and parafoveal locations. We present clear evidence of contrast perceptual learning at the behavioural level and show that these changes have neuronal correlates primarily in V4, rather than in V1. Learning was specific to stimulus location and spatial frequency, but was transferable across orientations; it took place to a limited degree under stimulus roving conditions, and could be either facilitated or impeded by the addition of flanker stimuli, depending on the subject. Upon removal of flankers, levels of psychometric and neurometric performance returned to their pre-flanker state.

In V4, learning-induced changes encompassed a shift in the point of neurometric equality and the semi-saturation constant (C_{50}) towards the trained contrast; a decrease in noise correlations across channels; and an increase in choice probability. In V1, enhancements in performance were characterised by an increase in spike discriminability; a shift in the point of neurometric equality and the C_{50} towards the trained contrast(s); and a widening in the range and a steepening of the contrast response function, during the early phase of training. Deteriorations in performance were accompanied by the reverse effects on V1 activity; furthermore, a general decrease in V1 firing rates occurred when training was carried out over an extended period of time, after performance had reached its peak.

Abbreviations

AD	analog to digital
AUROC	area under the ROC curve
BOLD	blood-oxygen-level-dependent
C_{50}	semi-saturation constant
CBC	Comparative Biology Centre
CD	contrast discrimination
CP	choice probability
cpd	cycles per degree
CRF	contrast response function
CSC	continuously-sampled channel
CSF	contrast sensitivity function
dva	degrees of visual angle
ECG	electrocardiogram
FDR	False discovery rate
FEF	frontal eye fields
FF	Fano factor
fMRI	functional magnetic resonance imaging
IPS	intraparietal sulcus
ISI	inter-stimulus interval
IT	inferior temporal cortex
JND	just-noticeable difference
LIP	lateral intraparietal area
M	mean
MBT	mixed-by-trial
MEA	multielectrode array
MEX	Matlab executable
MUA	multiunit activity
NHP	non-human primate
PL	perceptual learning
PO	preferred orientation
PROBMAT	probability matching of within-trial activity

PSE	point of subjective equality
PSTH	peristimulus time histogram
PNE	point of neurometric equality
RF	receptive field
ROC	receiver operating characteristic
RT	reaction time
SD	standard deviation
SEF	supplemental eye field
SEM	standard error of the mean
SF	spatial frequency
SNR	signal-to-noise ratio
SUA	single-unit activity
TvC	threshold versus contrast
V1	visual area 1
V2	visual area 2
V4	visual area 4

List of tables

List of figures

Chapter 1: Contrast discrimination task

1.1 Literature review

Organisation: this review starts by describing the phenomenon of perceptual learning (PL) and its general characteristics. It highlights important studies that have shed light on the mechanisms underlying PL, and presents three hypothetical models that explain how such learning might be implemented in the visual system. Finally, it concludes with a delineation of the goals of this project and summarises the key questions that will be addressed in the rest of this thesis.

1.1.1 What is perceptual learning?

Perceptual learning is a long-lasting improvement in the ability to make fine perceptual discriminations, achieved through practise, over many trials. Perceptual enhancements may persist for weeks, months, or even longer (Avi Karni & Sagi, 1993; Zhou et al., 2006), in contrast with the relatively short-lived changes seen during adaptation, sensitisation, and priming. The speed and extent of perceptual improvement depend on the nature of skills required. Tasks range from the complex (involving several perceptual dimensions and thus requiring the performance of discriminations at a more 'global' level) to the simple (involving only one feature dimension and thus likely to be mediated by specialised perceptual machinery). On the whole, 'global' tasks seem to be learnt more easily and result in greater improvement than 'simple' tasks (for a review, see Fine and Jacobs (2002)). Furthermore, transfer of learning, from a highly familiar task to a new one, appears to occur more readily from complex to simple activities, than vice versa (Ahissar & Hochstein, 1993; Fahle, 2005).

Studies conducted in the visual modality have reported enhancements in the discrimination of stimulus features, such as the orientation of lines and gratings (Ahissar & Hochstein, 1993; Furmanski, Schluppeck, & Engel, 2004; Ghose, Yang, & Maunsell, 2002; Kahnt, Grueschow, Speck, & Haynes, 2011; Matthews, Liu, Geesaman, & Qian, 1999; Raiguel, 2006; Schoups, Vogels, Qian, & Orban, 2001; Yang & Maunsell, 2004; T. Zhang, Xiao, Klein, Levi, & Yu, 2010; Zivari Adab & Vogels, 2011), the degree of separation or alignment between stimuli in vernier and bisection tasks (Crist, Li, &

Gilbert, 2001; Levi, 2005; Levi & Polat, 1996; Levi, Polat, & Hu, 1997; R. Li, Klein, & Levi, 2008; R. Li & Levi, 2004; R. Li, Provost, & Levi, 2007; R. Li, Young, Hoenig, & Levi, 2005; W. Li, Piëch, & Gilbert, 2004; Parkosadze, Otto, Malania, Kezeli, & Herzog, 2008; Xiao et al., 2008), the direction and speed of moving stimuli (Gu et al., 2011; Law & Gold, 2008; Liu & Vaina, 1998; Saffell & Matthews, 2003; Seitz, Nanez, Holloway, Koyama, & Watanabe, 2005; Zanker, 1999), the segregation of elements based on texture (A. Karni & Sagi, 1991; Schwartz, Maquet, & Frith, 2002; Yotsumoto, Watanabe, & Sasaki, 2008), and the depth disparity of perceptually-misaligned objects (Fendick & Westheimer, 1983; Ramachandran & Braddick, 1973; Westheimer, 1996).

Improvements are often reported as being closely dependent upon the specific stimuli to which subjects are exposed, and are not readily transferable to non-trained stimulus parameters. For instance, when stimuli consist of a series of gratings that differ subtly across various parameters, improvements in the identification and discrimination of stimuli are highly specific to the orientation (Ahissar & Hochstein, 1993; Dorais & Sagi, 1997; Ghose et al., 2002; Raiguel, 2006; Schoups et al., 2001; Shapley, 2003), spatial frequency (Sowden, Rose, & Davies, 2002), contrast (Crist et al., 2001; Polley, 2006; Shapley, 2003), size (Ahissar & Hochstein, 1993), and visual field location (Schoups et al., 2001; Sowden et al., 2002), of those used during training sessions (for reviews, see Fahle (2005), Dosher and Lu (2004), Lu, Hua, Huang, Zhou, and Dosher (2011) and Gilbert, Sigman, and Crist (2001)).

1.1.2 Contrast discrimination in human psychophysics studies

Visual stimulus contrast is sometimes viewed as a special case- the discrimination of objects with low luminance contrast is a daily component of the visual diet, and the contrast discrimination faculties of adult primates are typically believed to have reached maximum levels of performance during normal development. Although a substantial body of clinical work has documented prolonged, marked improvement in contrast detection amongst amblyopic patients as a result of training (Chen, Chen, Fu, Chien, & Lu, 2008; Chung, Li, & Levi, 2006, 2008; Huang, Zhou, & Lu, 2008; Polat, Ma-Naim, Belkin, & Sagi, 2004; Polat, Ma-Naim, & Spierer, 2009; Zhou et al., 2006), the scope for learning in humans with normal vision was thought to be limited. This view was supported by early studies in healthy humans where improvements in contrast

discrimination (CD) tasks were minimal- or at least, highly specific to the contrast levels used during training (Adini, Sagi, & Tsodyks, 2002). In one extreme example (documented by Tsodyks, Adini, and Sagi (2004)), practise on a CD task for as many as 40 training sessions failed to yield significant improvements in contrast thresholds.

However, findings by Yu et al. (2004) suggested that improvements in CD were, in fact, possible, and could be achieved in most subjects by carrying out the training regimen for an extended period. Adini, Wilkonsky, Haspel, Tsodyks, and Sagi (2004) repeated their experiment with new subjects (Experiment 5 of Adini et al. (2004)), and on that occasion, they reported significant learning effects.

Improvements in contrast sensitivity as a result of training have now been convincingly documented in humans with normal vision (Adini et al., 2004; Kuai, Zhang, Klein, Levi, & Yu, 2005; Xiao et al., 2008; Yu, Klein, & Levi, 2004; J.-Y. Zhang et al., 2008), providing a detailed picture of the circumstances under which learning takes place. The amount of learning that occurs depends on numerous factors- the abilities of individual subjects; their learning speed; and the particular tasks that they carry out. Where limited improvement or minimal transfer of learning is reported, this could be because previous training sessions did not provide subjects with 'sufficient practice' (Yu et al., 2004).

In animal studies, training can be carried over much longer periods, spanning several months. Thus, the following sections present findings from electrophysiology experiments on perceptual learning in animals, most of which were conducted on non-human primates (NHPs).

1.1.3 Electrophysiological signatures of perceptual learning

To identify the neuronal changes that accompany behavioural improvements in perception, many studies use electrophysiological single-unit recordings. To date, examinations of learning-induced changes in activity have been made using one of several methods of comparison:

1. Recordings could be compared between trained and untrained animals (e.g. Hua et al. (2010)).

2. Recordings could be taken from the same animal, but from different hemispheres (where one hemisphere corresponds to the trained retinotopic location, and the other to the 'untrained' location, e.g. Ghose et al. (2002), Crist et al. (2001), Yang and Maunsell (2004), Raiguel (2006)).
3. Recordings could be taken from the same animal and hemisphere, but from different retinotopic sites, e.g. Schoups et al. (2001).
4. Recordings could be taken from the same cortical region in the same animal, at different time points (over the course of training, e.g. Law and Gold (2008) and Zivari Adab and Vogels (2011)).
5. Recordings could be taken from the same cortical region in the same animal, but in response to familiar or unfamiliar stimuli (familiar stimuli are those used during training, whereas unfamiliar stimuli are either completely novel to the animal, e.g. Rainer et al. (2004), or are behaviourally unimportant, e.g. Schoups et al. (2001)).

Contrast-dependent changes in V1 were observed in anaesthetised cats, after subjects underwent training on an orientation discrimination task for over a month. Hua et al. (2010) found that after cats underwent training on an orientation task, the contrast thresholds and C_{50} contrast sensitivities of their V1 neurons were significantly better than those recorded from untrained cats. An examination of the contrast sensitivity function (CSF) revealed that learning was associated with increased contrast gain and a leftward shift of the contrast response function (CRF). This study was unusual because the cats had been trained not on a CD task, but on an orientation discrimination task. This indicated that the explicit direction of attention to the feature of interest was not necessary for PL. Another noteworthy aspect of this experiment, as pointed out by Lu et al. (2011), was that recordings were conducted under anaesthesia, and V1 may have received weaker top-down modulatory signals than if the animals had been awake. The neurometric improvements observed in V1 were thus likely to have been triggered primarily by localised changes in activity, than to have been driven by attention-based mechanisms in higher cortical regions.

In NHPs, however, much less is known about the capacity for improvement on a CD task. For the reasons mentioned earlier, the innate capacity for fine contrast

discrimination is thought to be particularly well-developed, and training-induced improvement is not guaranteed. Thus, previous neuronal recording studies on PL in monkeys have focused on orientation discrimination (Raiguel, 2006; Schoups et al., 2001; Yang & Maunsell, 2004; Zivari Adab & Vogels, 2011), motion discrimination (Law & Gold, 2008), or line bisection tasks (Crist et al., 2001; W. Li et al., 2004), rather than contrast discrimination. The following section summarises the current state of knowledge about neuronal correlates of PL in the primate visual cortex, and describes hypotheses that may be extrapolated to the contrast domain.

1.1.4 Models of perceptual learning

This section of the review frames the on-going debate based on three predominant theoretical models of PL, and describes evidence for each (Figure 1 provides a simple illustration of the sites of plasticity along the visual hierarchy, as proposed by the three models).

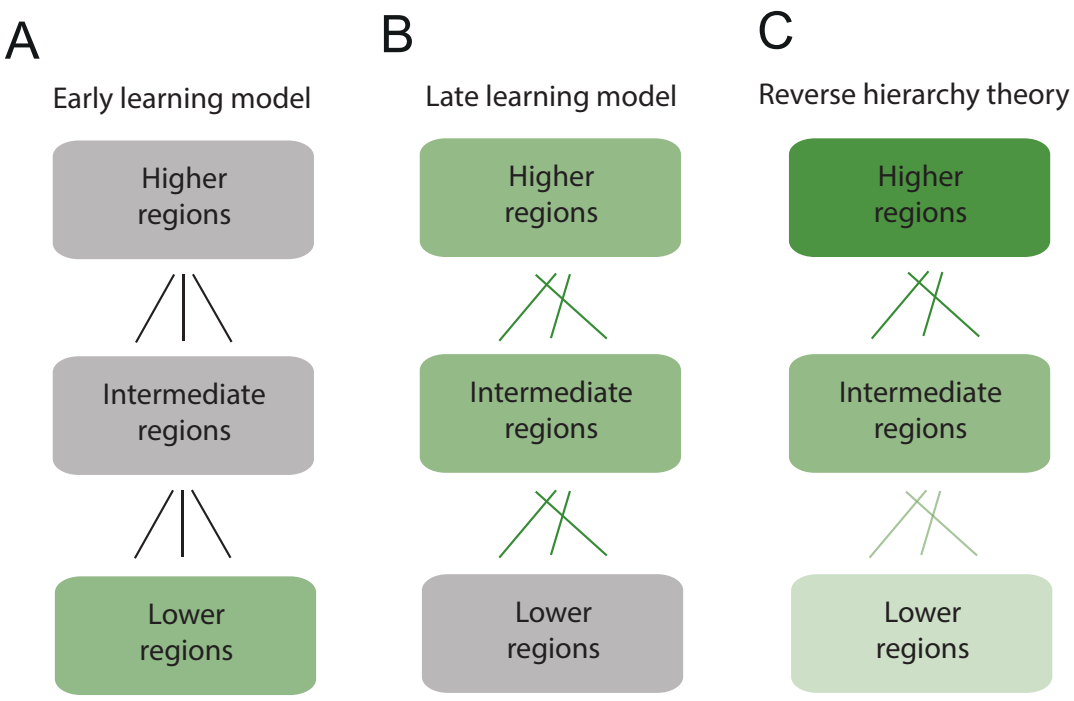


Figure 1. Schematic diagram of proposed sites of plasticity during perceptual learning, according to each of the three models. Grey boxes: no changes occur in these regions; green boxes: changes do occur. Grey lines between boxes: no changes in connectivity occur between these regions; green lines between boxes: changes in connectivity do occur. The early learning model (A) suggests that changes occur in regions such as V1

and V2; the late learning model (B) suggests that they occur in V4, TEO, IT and LIP; while the RHT (C) proposes that changes propagate from higher to lower areas.

At opposite ends of the spectrum lie two opposing theories: the 'early learning' model proposes that lower levels of the visual hierarchy undergo the most change, whereas the 'late learning' model argues that adjustments occur predominately within higher regions. The third model, termed the 'reverse hierarchy theory of learning,' provides a more unified account of events, proposing that changes are initially implemented at higher regions, and then occur at lower regions. Electrophysiology studies make it possible to identify changes in neuronal activity as learning progresses and to verify the strength of such claims.

1.1.4.1 Early learning model

This model predicts that the specificity observed in PL (e.g. to parameters such as the spatial frequency and spatial location of stimuli used during training) occurs primarily as a result of plasticity and reorganisation at 'lower' levels of the visual system, e.g. V1 and V2, rather than at higher levels such as inferotemporal (IT) cortex. This is because the tuning properties of neurons in lower regions (such as small receptive fields; responsiveness to a narrow range of stimulus parameters; and highly precise retinotopic mapping (Gilbert et al., 2001)) make them highly suitable for the processing of stimuli at the level of specificity that is often required. As one ascends the visual hierarchy, input from lower-level sources converges upon higher cortical areas, resulting in progressively broader tuning properties of neurons at each level (yielding larger receptive fields (RFs), for example).

The following paragraphs describe findings from several electrophysiology studies that are used in support of this lower-level mode of learning (Crist et al., 2001; W. Li et al., 2004; Schoups et al., 2001).

Schoups et al. (2001) trained monkeys to perform orientation discrimination tasks and compared V1 responses at trained and untrained locations. Performance improved markedly with training and was specific to stimulus location and orientation. The researchers observed shifts in preferred orientation (PO)- not across the recorded population as a whole, but rather, amongst cells with tuning preferences that rendered

them well-equipped to signal subtle differences in trained orientations. Specifically, they reported increases in the slope of the orientation tuning curve at the trained orientation for this select group of cells.

Crist et al. (2001) carried out an examination of V1 responses in macaques during training on a visual bisection task. The amount of modulation observed was compared between bisection and passive fixation trials, in trained and untrained locations. Responses were significantly modulated in the trained hemisphere, during presentation of trained stimuli. While basic RF size, cortical magnification, and orientation tuning properties showed no changes after training, the researchers observed task-dependent enhancements in the degree of modulation (whether excitatory or inhibitory) in trained animals, which occurred specifically for stimuli that were used during training. Furthermore, the size of these effects depended on the distance between the bar stimuli used during the task. Crist et al. postulated that these effects might have arisen through local changes in the balance of facilitatory and inhibitory horizontal inputs to V1 neurons, which might themselves have been modulated by feedback from higher-order areas.

Thus far, these tasks could be described as relatively 'simple,' as comparisons were made between levels of activity that were elicited during performance of a particular task, and those obtained under passive viewing conditions. To investigate the potential involvement of low-level regions in more complex tasks, W. Li et al. (2004) took the logical next step of asking whether the modulations observed by Crist et al. were present when subjects attended to different attributes of identical stimuli. The researchers recorded from macaque V1 neurons while subjects were presented with five line stimuli, and performed either a bisection or a vernier task. The authors observed task-dependent modulation of V1 responses, in the form of a steepening of 'offset tuning curves' (responses as a function of the degree of separation between lines). They proposed that V1 was a principal site of learning for two key reasons. Firstly, the task required high resolving power (a few arc minutes of visual angle)- a role that is compatible with V1. Secondly, task-dependent modulations (depending on whether a bisection or vernier task was performed) appeared early on in the V1 response- presumably too rapidly to have been due to feedback from higher areas.

A human fMRI study by Jehee, Ling, Swisher, van Bergen, and Tong (2012) found a positive relationship between behavioural improvements on an orientation discrimination task, and improvements in the signal-dependent discriminability of individual voxel responses in regions of interest. These effects were observed for voxels in trained V1 locations, but not in higher visual areas or untrained V1 locations, thus providing support for the early learning model.

If training on complex cognitive tasks indeed triggers changes in V1 or V2, as the early learning model claims, then how might these regions be targeted as candidates for plasticity? Top-down attention has been hypothesised to modulate activity across multiple areas in the visual hierarchy, and restrict the site of long-lasting modifications to lower areas during perceptual learning (Hochstein & Ahissar, 2002). As the neuronal response in each region evolves with time, lower-level processing areas which possess higher specificity in stimulus representation may become increasingly targeted by later-occurring, narrowly-focused components of attention, and it is the changes at these areas, according to the early learning model, that give rise to PL.

At the cellular level, learning-induced changes in an orientation discrimination task, for example, might be implemented via axon collateral interactions between V1 superficial pyramidal cells. These long-ranging horizontal connections extend over several millimetres (Sceniak, 2001), allowing communication between units with similar orientation preferences (Yoshimura, Dantzker, & Callaway, 2005). PL might occur through selective modulation of subsets of these horizontal connections, resulting in highly stimulus-specific improvements in performance (Crist et al., 2001).

1.1.4.2 Late learning model

Conversely, the late learning model suggests that changes at the neuronal level occur further up in the visual hierarchy. It proposes alternative explanations for the specificity of PL: for example, plasticity may be mediated by cells in higher level regions which remain narrowly-tuned to stimulus properties; or, modifications in the readout of signals by higher levels regions may occur without any significant involvement of V1 (Ghose et al., 2002). A number of studies have observed alterations in areas such as TEO, IT, and LIP, while others report that the changes which took place

at lower visual areas were unable to fully account for the degree of behavioural improvement attained through training (Ghose et al., 2002; Law & Gold, 2008; Mollon & Danilova, 1996; Raiguel, 2006; Rainer, Lee, & Logothetis, 2004; Williford, 2006; Yang & Maunsell, 2004).

According to the late learning model, enhancements in discrimination arise through a process of reweighting- when changes in connectivity occur between neurons from lower and higher visual areas (Yotsumoto & Watanabe, 2008). Neurons in higher regions are thought to selectively gate the inputs from lower regions, thereby fine-tuning the 'readout' of sensory information from lower areas.

In a study that is widely cited in support of this theory, Ghose et al. (2002) trained monkeys in an orientation discrimination task and found that learning-induced improvements were specific to trained orientations, but not to trained retinal locations. Furthermore, they observed small but significant decreases in the V1 population response to the trained orientation, at the trained location; however, these slight modifications were insufficient to account for the orientation specificity that was observed at the behavioural level. Overall, they found responses in both V1 and V2 to be extremely similar between trained and untrained regions. They therefore suggested that behavioural improvements arose from task-dependent and orientation-selective pooling of signals by higher areas.

Following this lack of evidence for extensive involvement of V1 or V2, researchers from the same lab turned their attention to V4, using the same task design (Yang & Maunsell, 2004). They then found that training was accompanied by decreases in tuning bandwidth and increases in response amplitude, particularly for neurons that had POs which differed from the trained orientation by ~ 45 degrees. Furthermore, the strength of correlations between neuronal firing rates and stimulus orientation (i.e. the discriminability of responses to various orientations) increased as training progressed.

Subsequently, Raiguel (2006) reported changes in V4, using the same paradigm as that used by Schoups et al. in their V1 study. They found that V4 neurons in the trained hemifield exhibited stronger responses and narrower orientation curves than those in the untrained hemifield. Changes were most obvious in the neurons which had

POs that differed by 25 – 65 degrees from the trained orientation, confirming the results previously obtained by Yang and Maunsell (2004) in V4, and mirroring the effects demonstrated in V1 by Schoups et al. (2001).

Zivari Adab and Vogels (2011) monitored V4 activity during a coarse orientation discrimination task, across a range of stimulus signal-to-noise ratios (SNRs). A comparison of single-unit activity between early and late recording sessions revealed an increase in response discriminability (measured as the area under the receiver operating characteristic curve, AUROC) and a decrease in variance (quantified by the Fano factor, *FF*), with learning. Furthermore, unlike the studies involving training on fine orientation discriminations, Zivari Adab and Vogels (2011) found that these effects were not restricted to the most informative subset of neurons, but were present across a broader spectrum of the sampled population. Their observations thus supported the idea that learning-dependent modulations of activity are tailored to the demands of the task, and that the coarser the discriminations required, the larger the pool of neurons that may potentially be affected.

Using a somewhat different paradigm from those described thus far, Rainer et al. (2004) examined V4 responses to novel and familiar stimuli in an object recognition task, defining learning in this case as that which occurs through prior exposure to a given stimulus. Learning was accompanied by higher levels of information in V4, when stimulus-evoked activity was compared between familiar and novel stimuli. In addition, the researchers found that when familiar images were 'degraded' through the addition of visual noise, this boosted the amount of information in the neuronal signal, as though the neurons were being specifically charged with the task of conveying maximal levels of disambiguating information under challenging conditions.

In another object recognition task, Baker, Behrmann, and Olson (2002) trained their monkeys to discriminate between a variety of tetrad stimuli. These stimuli were made up of four component batons, and the component that elicited the highest responses was termed the 'best' baton. Interestingly, the researchers did not find changes in absolute response strength between the best learned and unlearned stimuli; rather, they found enhancements in the discriminability of IT responses to component parts *within* learned stimuli, compared to that within unlearned stimuli. Furthermore,

these increases in neuronal selectivity were clearest when viewed at the population level, rather than at the level of individual neurons. These results differed from those seen previously in lower visual areas, where alterations were strongest for a select subpopulation of neurons.

Learning-induced modulations have also been found in LIP, a sensorimotor area, during training on a direction of motion task. Law and Gold (2008) observed the responses of neurons that were initially tuned for saccadic direction and showed no preferences for the direction of stimulus motion. Over the course of training, these neurons grew increasingly responsive, and their activity levels were increasingly well correlated with motion strength and viewing time. (The researchers also recorded from MT neurons during training on motion stimuli, but only observed increases in choice probability (CP), and not in motion sensitivity, in this area.)

These findings support the idea that medium-to-higher level cortical areas are actively involved in perceptual learning.

1.1.4.3 Reverse hierarchy theory of learning

The third model, the reverse hierarchy theory (RHT) of learning, incorporates elements from both late and early models, suggesting that changes occur throughout the visual hierarchy, but are overseen by high-level cognitive processes (Ahissar & Hochstein, 2004; Hochstein & Ahissar, 2002). It proposes that attention mechanisms 'alert' the cortex to behaviourally-relevant stimuli, and that a form of gating is carried out by neuromodulators that operate in task-relevant regions, enabling plasticity. Thus, it proposes that top-down mechanisms such as attention are responsible for selective alterations of appropriate neuronal populations.

This hypothesis is supported by the observation that naïve, untrained performers tend to show improvements in higher-level aspects of complex tasks, before acquiring the ability to make fine perceptual discriminations. The RHT proposes that when new performers first engage in a task, initial reorganisation occurs at higher cortical regions, and that this state of plasticity contributes to the acquisition of broad perceptual skills, which are transferable across a variety of related tasks. With continued practice, changes propagate downwards, towards lower-level neuronal populations in the visual

hierarchy. Gradually, areas that are responsible for making relatively fine perceptual distinctions become 'wired up' more efficiently.

Once expertise is acquired, the activation of higher-level volitional processes triggers off an automated cascade, where lower-level neuronal populations that are responsible for fine perceptual discriminations send output readily to higher regions. The initial reorganisation at higher levels in beginners would tend to hone broad perceptual skills that are transferable across a variety of stimulus parameters. Modifications that occur later on, on the other hand, occur primarily in lower regions, and are likely yield to higher specificity in discrimination skills and less generalisation across stimulus features.

1.1.5 Effects of attention on contrast response functions of visually-responsive neurons

Attention is known for its modulatory effects on visually-evoked responses to stimuli of various contrasts, characterised variously as a response gain in V4 (McAdams & Maunsell, 2000; Treue & Maunsell, 1999; Treue & Trujillo, 1999) and MT (Lee & Maunsell, 2010), a contrast gain in V4 (Carrasco, Ling, & Read, 2004; Reynolds, 2000) and MT (Martínez-Trujillo & Treue, 2002), as an additive model in V1 (Buracas & Boynton, 2007; Thiele, Pooresmaeili, Delicato, Herrero, & Roelfsema, 2009), or potentially any one of these possibilities, in V4 (Williford, 2006).

In the response-gain model, attention scales firing rates to a degree that is proportionate to the size of the response elicited in the absence of attention. Thus, the greater the baseline response to a given stimulus, the more strongly it is up-regulated by attention. In the contrast gain model, the saturation points of the CRF remain fixed, while the responses elicited by low-to-intermediate contrasts are boosted by attention, effectively shifting the CRF towards the left. In an additive model (also referred to as an 'activity model'), attention increases the response by a relatively fixed amount across a wide range of supra-threshold contrasts.

If PL-induced changes were mediated in part by attentional mechanisms, one would expect the effects of learning on neuronal responses to mimic those observed

during the engagement of attention. Thus, one might find a shift in the CRF towards the left, or upwards; it might affect a select range of stimulus contrasts, or operate across a broader range. In theory, learning might even be accompanied by effects that resemble the *disengagement* of attention, i.e. a rightward shift in the CRF, and/or a down-regulation of responses.

1.1.6 Goals of the contrast discrimination task

In summary, several cortical regions are known to be involved in the learning of a variety of perceptual tasks, but the biological underpinnings of CD learning remain relatively unknown and the exact locations of plasticity are under dispute, making CD a promising domain for further study. We do not yet have a clear understanding of the contributions of each region at specific points in time, neither a coherent picture of how various regions interact to yield perceptual gains. Furthermore, CD is known to be a perceptually demanding task (it was only within the last decade that human subjects were definitively shown to be capable of substantial improvement)- this raised the question of whether similar gains would be possible in macaque subjects, and simultaneously ensured that any changes in fine discrimination, if present, would take place over a prolonged period and could thus be monitored in close temporal detail.

Human psychophysics studies provide insights into the perceptual improvements that result from training on a CD task, while electrophysiology studies have identified changes at the neuronal level in the primate brain during training on an assortment of other visual tasks. A combination of the two bodies of literature thus offers a guide map for the examination of the neuronal underpinnings of perceptual learning in the contrast domain.

In the majority of electrophysiological studies described earlier, single-unit recordings were made using acute electrodes, and activity was recorded from a small number of neurons at a time. The exact location of the recording electrode changed from day to day, resulting in the sampling of different subpopulations of neurons across sessions. Ideally, recordings would be made from a stable subpopulation of neurons, across the entire training period, from the same animal, as this would reduce levels of

variability due to sampling differences across recording sessions, and provide stronger support for the argument that changes (if present) are indeed due to training.

A considerable advantage of using NHPs is that experiments can be conducted near-daily for weeks or months if necessary- an undertaking that would be infeasible in most human studies. With chronically-implanted multielectrode arrays (MEAs), it is possible to obtain multiunit recordings from a relatively stable pool of neurons over an extended period of time. Such arrays yield satisfactory signals from a large number of channels, and grids can remain fixed in place for years, with good signal quality throughout (Simeral, Kim, Black, Donoghue, & Hochberg, 2011).

The aim of the current study was thus to record from macaque V1 and V4 using chronically-implanted MEAs, to monitor the behavioural effects of training on contrast discrimination abilities, and to investigate whether concurrent changes in spiking activity occurred in these two regions.

1.1.6.1 Psychophysics/ behavioural questions

- With training, do adult macaque subjects show improvements in fine contrast discrimination?
- If so, to what degree is this possible, and what is the time course of learning?
- Are improvements specific to stimulus properties such as location, orientation, and spatial frequency?
- What signatures of the psychophysical functions change with learning?

1.1.6.2 Neurophysiological questions

- Are improvements accompanied by changes at intermediate and low-level regions of the visual cortex (V4 and V1)?
- What is the nature of these changes (e.g. alterations of firing rate, spike variance, and tuning properties)?
- What are the potential readout mechanisms employed by the system to mediate behaviour?
- Are the changes seen at neuronal level able to account for those seen at the behavioural level?

1.2 Neuronal recording methods

1.2.1 Data collection

All procedures were carried out in accordance with the European Communities Council Directive RL 2010/63/EC, the US National Institutes of Health Guidelines for the Care and Use of Animals for Experimental Procedures, and the UK Animals Scientific Procedures Act. Two male macaque monkeys (5 – 14 years of age) were used in this study.

1.2.1.1 Head post implantation

An initial surgical operation was performed under sterile conditions, in which a custom-made head post (Peek, Tecapeek) was embedded into a dental acrylic head stage. Details of surgical procedures and post-operative care have been published elsewhere (Thiele, Delicato, Roberts, & Gieselmann, 2006).

1.2.1.2 General training

Initially, monkeys were habituated to perform a delayed match-to-sample task, in which they compared the colour of a circle stimulus with that of succeeding circle stimuli, while maintaining fixation on a central target. When a target stimulus appeared (a circle of a matching colour), subjects were required to release a touch bar in order to receive a fluid reward. Eye position was monitored using an infrared video tracking system (Dalsa CCD camera [model SIM-0002] and eye-tracking software from Thomas Recording ET-49 [Version 1.2.8]). This allowed subjects to familiarise themselves with the experimental setup and the timing structure of the task; this task was otherwise unrelated to the CD experiment described in this thesis.

1.2.1.3 Electrode array implantation

During surgery, animals were sedated with ketamine, and general anaesthesia was maintained using isoflurane following endotracheal intubation. Heart rate, respiratory rate, blood pressure, ECG, O_2 saturation, expiratory CO_2, and skin and rectal

body temperature were monitored continuously during the operation. Fluids and antibiotics were administered intravenously.

The animals were placed in a stereotaxic head holder and the skull overlying the occipital and posterior temporal cortices was exposed. A craniotomy was made to remove the bone overlying V1, V2, and dorsal V4, using a pneumatic drill. The bone was kept in sterile 0.9% NaCl for refitting at the end of the surgery. The dura was opened up to allow access to regions V4 and V1. Microelectrode chronic Utah arrays, attached to a CerePort™ base (Blackrock® Microsystems, connection dimensions of 16.5 mm [height] × 19 mm [base diameter] × 11 mm [body diameter]), were implanted under sterile conditions in the cortex, using a Blackrock microarray inserter. In monkey 1, two 4 × 5 grids of microelectrodes were implanted in area V4, and one 5 × 5 grid was implanted in V1; in monkey 2, a 5 × 5 grid was implanted in V4, and a 5 × 5 grid in V1. Electrodes were 1 mm in length, and their tips reached depths of up to 1 mm, for grids in both striate and extrastriate cortex. For grids that were embedded in striate cortex, recordings were thus estimated to arise predominantly from layer 3 neurons. Wire bundles were held in place with biologically-compatible glue (histoacrylic), and the connector (CerePort™) was secured to the skull with titanium bone screws. In both animals, the titanium screws were rejected by the bone within ~ 6 – 10 weeks following the implant, so a dental acrylic bridge was built to fuse the base of the connector to the existing head stage, during a subsequent surgical operation.

1.2.1.4 Data recording

Once animals had fully recovered, RFs were mapped using a reverse correlation procedure (DeAngelis, Freeman, & Ohzawa, 1994; Gieselmann & Thiele, 2008), for each recording channel. (Data processing and RF mapping procedures are described in detail in Appendix A: Artifact removal from neuronal data, on page 213, and in Appendix C: Characterisation of neuronal tuning properties, on page 237, respectively.) The aggregate RF for the V4 arrays was centred at visual coordinates of approximately (-5, -16) in each of the monkeys, while the locations of the V1 RFs differed slightly between the two monkeys (Figure 2 shows RF locations in monkey 1; Figure 3 and Figure 4 show RFs in monkey 2). In monkey 1, the V1 electrodes were positioned at a cortical location corresponding to 4.6° from the centre of vision, while in monkey 2, the

RFs were positioned much closer to the fovea, at 1.5°. In macaques, the fovea encompasses approximately the central two degrees of vision, i.e. up to one degree from the centre (Hanazono, Tsunoda, Kazato, Suzuki, & Tanifuji, 2012); thus in both animals, the V1 receptive fields were located outside the fovea, but inside the parafoveal region.

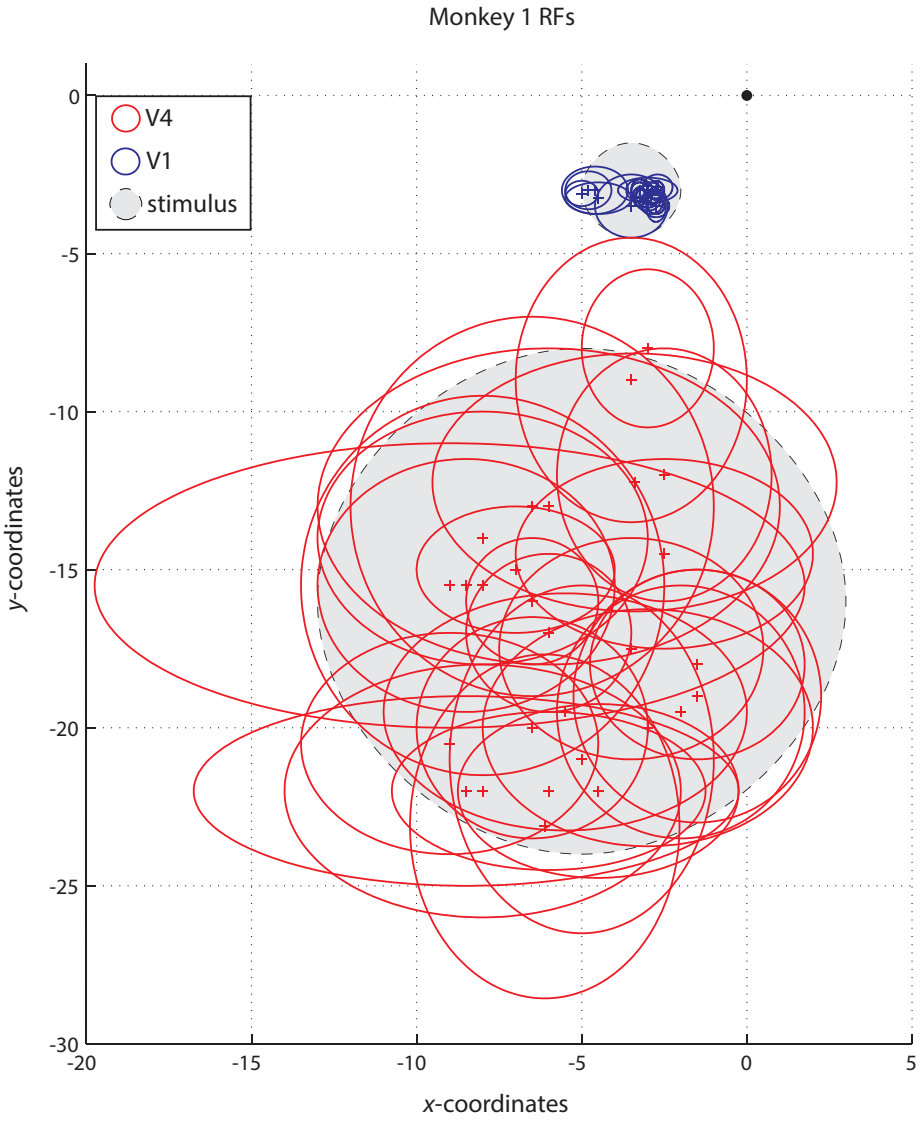


Figure 2. Receptive field and stimulus locations in monkey 1. The fixation spot is marked by the small black circle at visual coordinates of (0,0). Ellipses depict neuronal RFs of V4 (red) and V1 (blue) channels. Grey circles indicate stimulus locations used in the experiments (described in detail in the section, 'Stages of training on the main contrast discrimination task,' page 21).

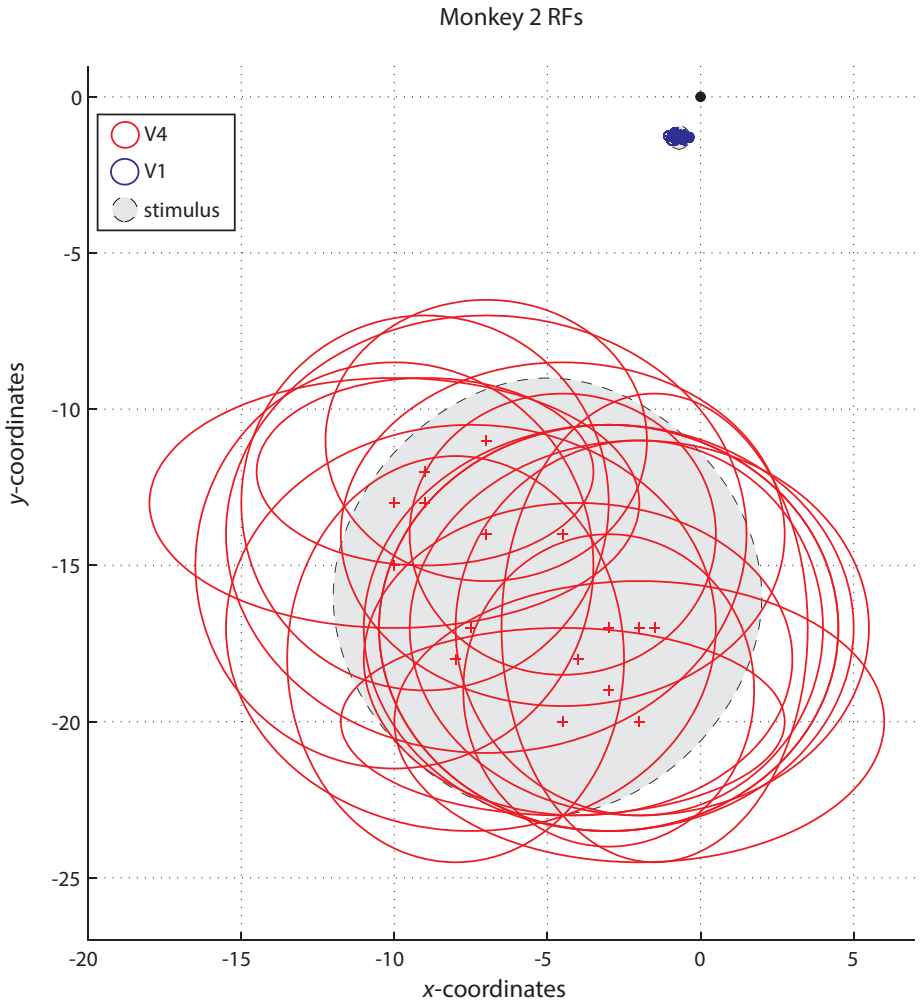


Figure 3. Receptive field and stimulus locations in monkey 2. The fixation spot is marked by the small black circle at visual coordinates of (0,0). Ellipses depict neuronal RFs of V4 (red) and V1 (blue) channels. Grey circles indicate stimulus locations used in the experiments. Refer to Figure 4 for a zoomed-in view of the V1 RFs.

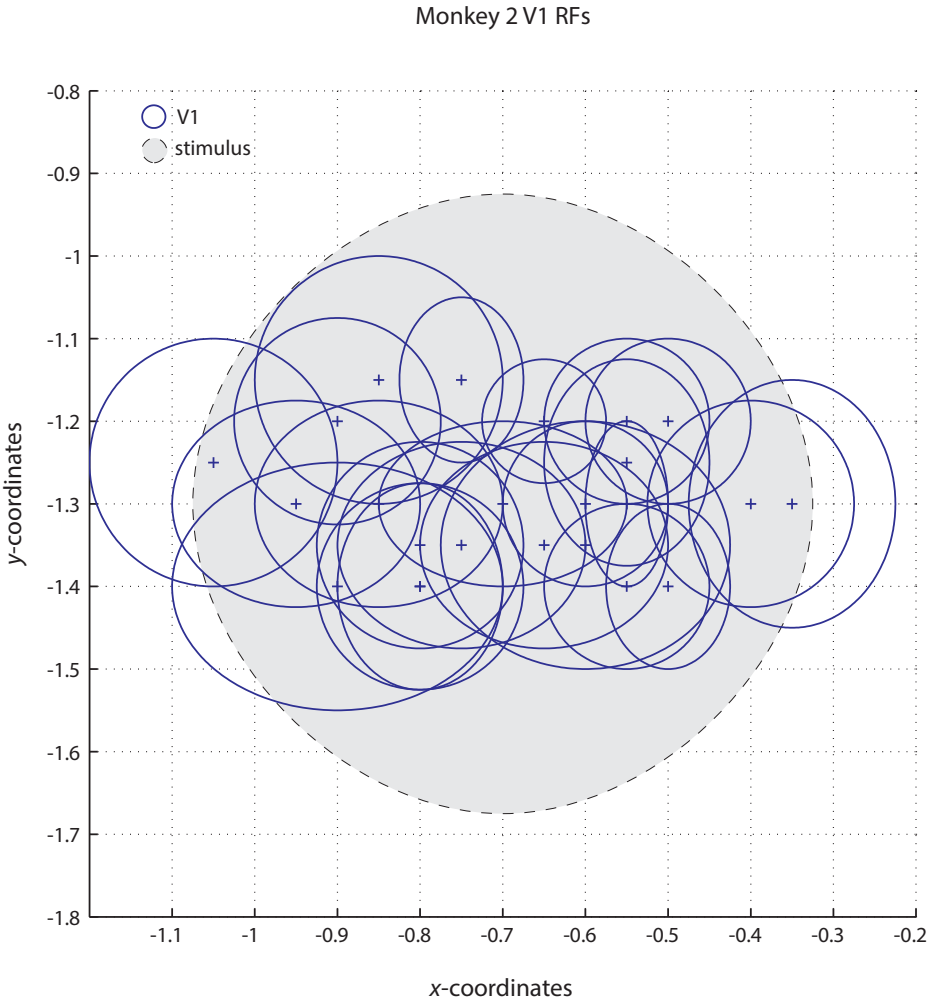


Figure 4. Zoomed-in view of V1 RFs in monkey 2.

Note that for presentations of small mapping stimuli (e.g. 0.1 dva in diameter), the question arose as to whether microsaccades may have caused slight deviations from the actual position of the stimuli in retinal coordinates. If this were the case, then the size of RFs may have been slightly over-estimated. However, this would not have affected our results in the CD task, as large-sized stimuli were intentionally chosen for the PL task, such that the stimuli filled and extended beyond the measured neuronal RFs, in the vast majority of cases. It is, however, unlikely that the size of the RFs was overestimated by a large amount, as the sizes of the V1 RFs reported here are well within the range of those reported for V1 recordings in anaesthetised and paralyzed macaques, for similar eccentricities, i.e. preparations in which eye movements are virtually absent.

1.3 Psychophysics methods

1.3.1 Stimuli

Stimulus presentation was controlled using Cortex software (Laboratory of Neuropsychology, National Institute of Mental Health, http://dally.nimh.nih.gov/index.html) on a computer with an Intel® Core™ i3-540 processor. Stimuli were displayed at a viewing distance of 0.54 m, on a 25" Sony Trinitron CRT monitor with a resolution of 1280 by 1024 pixels, yielding a resolution of 31.5 pixels/degree of visual angle (dva). The monitor refresh rate was 85 Hz for monkey 1, and 75 Hz for monkey 2. The output of the red and green guns was combined using a Pelli-Zhang video attenuator, yielding a luminance resolution of 12 bits/pixel, allowing the presentation of contrasts that were well below CD thresholds (Pelli, 1991). A gamma correction was used to linearize the monitor output.

1.3.2 Contrast discrimination task paradigm

Monkeys were engaged in a CD task, in which the presentation of a sample stimulus was followed by that of a test stimulus. They had to decide whether the contrast of the test stimulus was higher or lower than that of the sample stimulus (see Figure 5 for an illustration of the task). The task paradigm was based on that commonly used in the human psychophysics literature (as described in the section, 'Contrast discrimination in human psychophysics studies,' on page 2), in order to make our results as comparable to those of previous studies as possible. The delayed match-to-sample design used in the human studies was well-suited to our needs, as it ensured that the subjects were presented with physical reference stimuli on each trial, and were thus able to learn the task contingencies fairly easily, via the delivery of reward feedback; this was an important requirement as our macaques could not be explicitly instructed on how to perform the task, unlike humans. It also allowed the detailed study of the effects of roving stimuli.

The task involved the discrimination of contrasts that varied over a substantial range- some discriminations were highly challenging, while others less so. This was done to ensure that the learning of fine contrast discriminations took place over a

prolonged period of time, to allow continuous monitoring of improvement over the training period.

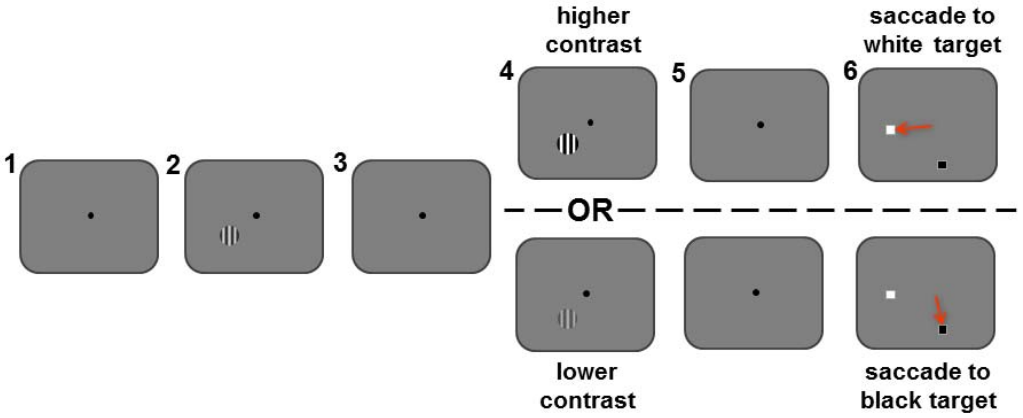


Figure 5. Illustration of the contrast discrimination task. 1) The monkeys were required to fixate upon a central spot, to initiate the trial. 2) While maintaining fixation, a sample stimulus of 30% contrast (either a Gabor patch or a sinusoidal grating) was presented for 512 ms in the lower left visual field. 3) Presentation of the sample stimulus was followed by an interval lasting 512 ms (except during training at the V4 location for monkey 1, where the interval lasted for a random duration of 512 to 1024 ms). 4) Next, the test stimulus (another Gabor patch or sinusoidal grating which could be of higher or lower contrast than the sample), was presented for 512 ms, 5) followed by a second interval of 400 ms. 6) Two target stimuli appeared to the left and right of the location at which the sample and test had previously been presented; the fixation spot changed colour from black to grey, signalling that the animals were allowed to make a saccade to their chosen target. If the test was of a higher contrast (e.g. 32%) than the sample (always 30%), the monkeys had to saccade to the white target; otherwise, if the test stimulus was of a lower contrast (e.g. 28%), they had to saccade to the black target. The red arrows in the figure indicate the direction of saccadic motion for illustrative purposes only; they did not appear onscreen.

1.3.3 Stages of training on the main contrast discrimination task

Training was carried out in three distinct phases. During the first and third phase, stimuli were positioned at a peripheral location in the visual field, corresponding to the receptive field (RF) location covered by electrodes in V4 (the 'V4 location'), and during the second phase, stimuli were positioned at the position corresponding to the location of the V1 electrodes (the 'V1 location'). Properties of the stimuli used throughout each stage of training are listed in Table 1.

1.3.3.1 Stage 1: Training with Gabor stimuli at the V4 location

Subjects performed the task with a Gabor stimulus, for several weeks (monkey 1: 30 sessions, spanning a period of 8 weeks; monkey 2: 26 sessions, spanning 6 weeks), until their performance reached a plateau. The sample stimulus had a contrast of 30%, while the test stimulus was presented at one of 14 possible contrasts [10, 15, 20, 25, 27, 28, 29, 31, 32, 33, 35, 40, 50, and 60%].

Property	Monkey 1			Monkey 2		
	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3
No. of sessions	30	17	5	26	22	5
Location	peripheral (V4)	parafoveal (V1)	peripheral (V4)	peripheral (V4)	parafoveal (V1)	peripheral (V4)
Coordinates of centre (dva)	(-5, -16)	(-3.5, -3)	(-5, -16)	(-5, -16)	(-0.7, -1.3)	(-5, -16)
Size (dva)	16	3	16	14	0.75	14
SF (cpd)	2	2	2	2	4	2
Orientation	vertical for all sessions but the last	vertical	vertical	vertical for all sessions but the last	vertical	vertical
Stimulus type	Gabor	sinusoidal grating	sinusoidal grating	Gabor	sinusoidal grating	sinusoidal grating

Table 1. Stimulus parameters used at each stage of contrast discrimination training.

At the end of training with a Gabor stimulus at the V4 location, we carried out an additional session during which the Gabor stimuli were horizontally, rather than vertically, oriented. This was to determine whether perceptual improvements would transfer to stimuli of an orthogonal orientation.

1.3.3.2 Stage 2: Training with sinusoidal grating stimuli at the V1 location

Following training at the V4 location, monkeys were trained to discriminate contrasts at the V1 location. The stimulus diameter was reduced from 16 to 3 dva in monkey 1 and from 14 to 0.75 dva in monkey 2. The sample stimulus had a contrast of

30%, while the test stimulus was presented at one of fourteen possible contrasts [5, 10, 15, 20, 22, 25, 28, 32, 35, 40, 45, 50, 60, and 90%].

In addition, a sinusoidal grating stimulus was used instead of a Gabor. This was because the perceived size of a Gabor changes with its peak contrast, such that a low-contrast Gabor seems smaller than a high-contrast one (Foley & Legge, 1981; Fredericksen, Bex, & Verstraten, 1997; Polat, 1999). Data were collected over 4-6 weeks (monkey 1: 17 sessions; monkey 2: 22 sessions).

1.3.3.3 Stage 3: Training with sinusoidal grating stimuli at the V4 location

To examine the effects of apparent size on task performance, we carried out a control experiment at the V4 location, in which we used sinusoidal grating stimuli instead of Gabor patches. This control was carried out for 5 sessions (1 week) for each of the subjects. As with the training carried out in Stage 1, the sample stimulus had a contrast of 30%, while the test stimulus was presented at one of fourteen possible contrasts [10, 15, 20, 25, 27, 28, 29, 31, 32, 33, 35, 40, 50, and 60%].

1.3.4 Measures of perceptual learning

To investigate the effects of perceptual learning, several metrics of performance were monitored over the course of training: the proportion of correct responses made by the subjects; the slope and the point of subjective equality of the psychometric function; the psychometric threshold (defined as the test contrast at which performance was at 81.6%); and the rate of learning for different contrasts.

The proportion of trials in which subjects made correct responses was calculated for each test contrast condition, yielding fourteen values of the contrast-dependent proportion of correct trials ('$P_{condition}$') per session. The average performance for each session ('$P_{correct}$') was simply the mean across these fourteen values of $P_{condition}$ and provided a broad overview of the subjects' daily performance across test contrast conditions.

From $P_{condition}$, we could calculate $P_{reporthigher}$, which was the proportion of trials in which subjects reported the test contrast as being higher than the sample contrast. A

Weibull function was fitted to values of $P_{reporthigher}$ using a maximum likelihood estimation method (Matlab, Mathworks), thus generating a psychometric curve for each session. The Weibull function was defined as

$$F_{\alpha,\beta}(x) = \delta - \gamma e^{-\left(\frac{x}{\alpha}\right)^{\beta}} \qquad \ldots \text{(Equation 1)}$$

where $F_{\alpha,\beta}(x)$ is the fitted value of $P_{reporthigher}$; x is the contrast of the test stimulus; γ is the range; δ is the maximum value; and α is the contrast at which $F_{\alpha,\beta}(x)$ reaches 63.2% of its maximum, which is occasionally used as a threshold measure when $F_{\alpha,\beta}(x)$ ranges from 0 to 1. In cases where $F_{\alpha,\beta}(x)$ does not range from 0 to 1 because γ and δ are freely varying parameters, it should not be considered a threshold, but simply the value that corresponds to $F_{\alpha,\beta}(x)$ when x = α. Lastly, β is the slope of the psychometric curve at x = α.

While the above equation yielded a slope for the contrast at x = α, this value did not necessarily provide an accurate representation of perceptual sensitivity at the most interesting and task-relevant part of the psychometric curve, i.e. close to contrasts of 30%. We therefore also determined the slope of the psychometric function at the point where the contrast was 30% (hereafter simply referred to as 'the slope'). This was calculated by finding the tangent to the fitted curve at the point x = 30% (depicted in Figure 6), according to the formula

$$slope = \frac{dF}{dx}\left[\delta - \gamma e^{-\left(\frac{30}{\alpha}\right)^{\beta}}\right] = 30^{\alpha-1}\alpha\gamma e^{-\left(\frac{30}{\alpha}\right)^{\beta}}\left(\frac{1}{\alpha}\right)^{\beta} \qquad \ldots \text{(Equation 2)}$$

Finally, we determined the point of subjective equality (PSE) of the psychometric function, which indicated the contrast at which the subject reported the test stimulus as being indistinguishable from the sample. The PSE was calculated by finding the contrast at which the value $F_{\alpha,\beta}(x)$ of the fitted function was equal to 0.5 (depicted in Figure 6). For a perfect observer, the value of the PSE would lie at exactly 30%; in our subjects, any deviation in the PSE from the value of 30% indicated a bias in their criterion level.

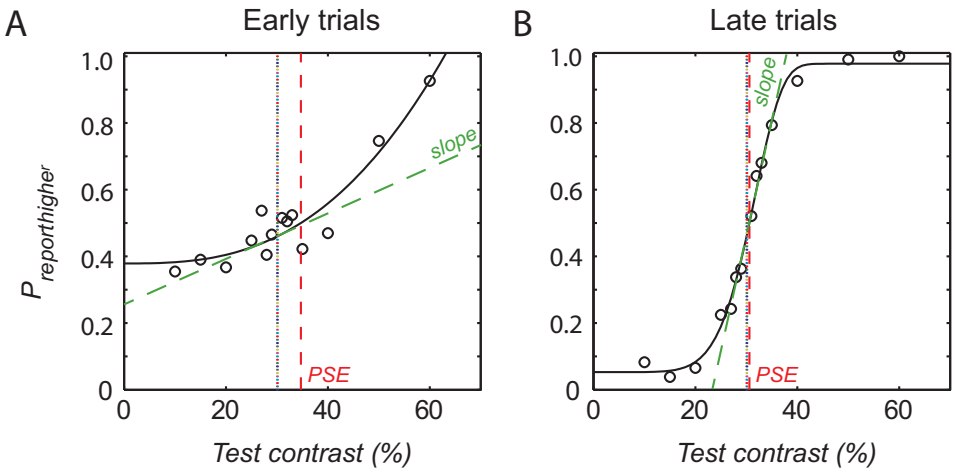


Figure 6. Illustration of hypothetical psychometric data, compared between early (A) and late (B) sessions. One would expect the slope to be relatively shallow for early sessions, and to grow progressively steeper with training. The PSE would also be expected to shift towards the value of the sample contrast (30%) over the course of training, regardless of its original location at the start of training.

To monitor changes in performance that occurred for each individual condition, values of $P_{reporthigher}$ were plotted against session number, as well as the running average (calculated across three sessions at a time).

1.3.5 Contrast thresholds

According to the threshold versus contrast (TvC) function in humans, for base contrasts above detection threshold, the size of the just-noticeable difference (JND) in luminance contrast between a stimulus and its increment depends on the absolute values of the contrasts being compared (Legge & Foley, 1980; Tsodyks et al., 2004; Wilson, 1980), in a manner similar to that predicted by the Weber-Fechner law (Fechner, 1860; Green & Swets, 1966; Weber, 1850). Accordingly, conditions with a lower-contrast test stimulus would be expected to yield smaller JNDs than conditions where the test was of higher contrast than the sample.

To address this possibility, we separated the conditions into two 'test contrast categories,' where the test contrast was (a) higher or (b) lower than the sample contrast (termed 'C_H' and 'C_L' conditions, respectively). These values were plotted against the absolute difference between the sample and test contrasts, and a Weibull curve was fit to the data in each category, according to the formula

$$G_{\alpha,\beta,\lambda}(|\Delta C|) = 0.5 + (0.5 - \lambda)(1 - e^{-\left(\frac{|\Delta C|}{\alpha}\right)^{\beta}}) \qquad \ldots \text{(Equation 3)}$$

where $G_{\alpha,\beta,\lambda}(|\Delta C|)$ is the fitted value of $P_{correct}$, with the bounds $0.5 \leq G_{\alpha,\beta,\lambda}(|\Delta C|) \leq \max(P_{correct})$; $|\Delta C|$ is the absolute difference between the sample and test contrasts; α is the threshold; β is the slope, with the bounds $0 \leq \beta \leq 5$; and λ is the proportion of erroneous responses for the condition which gave the highest value of $|\Delta C|$ during a given session (λ was set separately for each of the groups C_H and C_L). The psychophysical threshold was defined as the test contrast at which the subjects' performance would be at 81.6% correct (Green & Swets, 1966; Thiele, Dobkins, & Albright, 2000), yielding two thresholds, T_L (for conditions where the contrast of the test stimulus was lower); and T_H (for conditions where the contrast of the test stimulus was higher).

Inclusion of the parameter λ in equation 3 was based on the assumption that task performance depended on two distinct skills: 1) An understanding of the task contingencies (i.e. to comprehend that the basic requirement of the task was to make a comparison between the stimuli- a skill which could occur through associative learning and may depend on levels of attention), and 2) The ability to perform the task at a fine level (i.e. to make accurate discriminations in contrast). During early training sessions, learning would be expected to occur primarily at an associational level. Once subjects had learnt the underlying principles of the task, refinements in perception were then likely to proceed at a more specific level.

In order to distinguish between these two types of task learning, we assumed that engagement of the latter skill was essentially absent for the easiest task condition, due to the large difference in contrast between the stimuli. Changes in performance for this particular condition over the course of training would thus be attributable to improvements of contingency/associational relationships between the task stimuli and the reward, and poor performance for these conditions during later stages of training would likely be due to attentional lapses or eye movement errors. Thus, inclusion of this model parameter enabled the examination of fine contrast discrimination learning, that occurred independently of conceptual task learning and of daily or trial-wise fluctuations in attention (Law & Gold, 2008).

1.3.6 Reaction times

The monkeys' reaction time (RT) was defined as the time taken by the subjects to make a saccade to the target, from the moment that the fixation spot changed colour. A Pearson's correlation analysis was performed separately for RTs for correct and incorrect trials, to determine whether RTs changed over the course of training.

1.3.7 Corrections for multiple comparisons

For tests of significance that involved multiple comparisons, a False Discovery Rate (FDR) correction for α-levels was applied where appropriate, to reduce the likelihood of making either too many false positives or too many incorrect rejections (Benjamini & Hochberg, 1995). This procedure yielded a '*q*-value,' which acted as an FDR analogue to the *p*-value.

1.4 Behavioural results

1.4.1 Perceptual learning with stimuli at the V4 and V1 locations

The performance of the two subjects (monkeys 1 and 2) in the main contrast discrimination task was assessed over 52 and 53 sessions respectively. This was carried out in three stages (Stages 1 to 3), with stimuli positioned peripherally at the V4 location during the first and third stages, and parafoveally at the V1 location during the second stage (details were described in the methods section, on page 21).

1.4.1.1 Performance during trials with variable interval durations

For monkey 1, when stimuli were presented at the V4 location, the duration of the blank interval between the presentation of sample and test stimuli was a randomly chosen value from 512 to 1024 ms. To examine whether interval duration had any effect on the monkey's performance, trials were categorised into two groups, based on interval length (the first and last quarters of interval lengths). No significant main effect of trial duration was observed (three-way ANOVA, $F(3,819) = 2.03$, $p = .108$), neither was there an interaction between trial duration and the other factors (trial duration × test contrast: $F(39,819) = 0.93$, $p = .588$; trial duration × session: $F(84,819) = 0.85$, $p = .822$). Thus, data from Stage 1 for this subject were combined with the rest of the data for subsequent analyses.

1.4.1.2 Perceptual learning for individual test contrast conditions

To investigate whether learning rates differed between test contrast conditions, performance was plotted separately for each condition (Figure 7). The measure of performance used, $P_{reporthigher}$, was the proportion of trials in which the subject reported that the test contrast was higher than that of the sample. A visual inspection revealed that for the easier conditions, performance increased relatively quickly and reached a plateau within a few sessions, whereas for harder conditions, performance levels rose more gradually over a longer period of time.

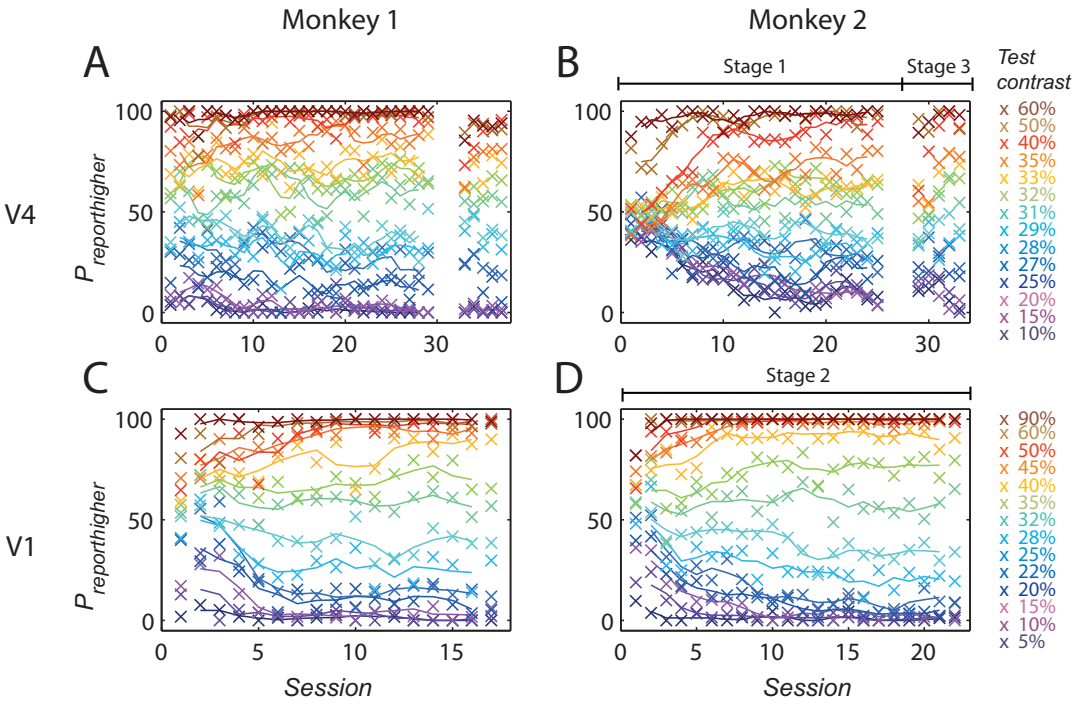


Figure 7. Proportion of trials during which the contrast of the test stimulus was reported to be higher than that of the sample, plotted against session, for each test contrast condition (coded by colour). A & B: V4 location (Stage 1, followed by five data points from Stage 3); C & D: V1 location (Stage 2). A & C: monkey 1; B & D: monkey 2. 'X' markers correspond to measured data, while lines depict the running average over three consecutive sessions, plotted for the middle session of the three. Changes in the value of λ with training (as described in the section, 'Psychometric thresholds for conditions with higher or lower test contrasts,' on page 31) are represented by an examination of changes in $P_{reporthigher}$ for the conditions with the highest (dark brown markers) and lowest (dark purple markers) test contrasts, respectively.

1.4.1.3 Perceptual learning across all fourteen test contrast conditions

Performance was assessed across all fourteen test contrast conditions, using three measures for each session: 1) the mean proportion of correct responses, 2) the slope of the psychometric curve at 30% contrast, and 3) the PSE of the psychometric curve (Figure 8).

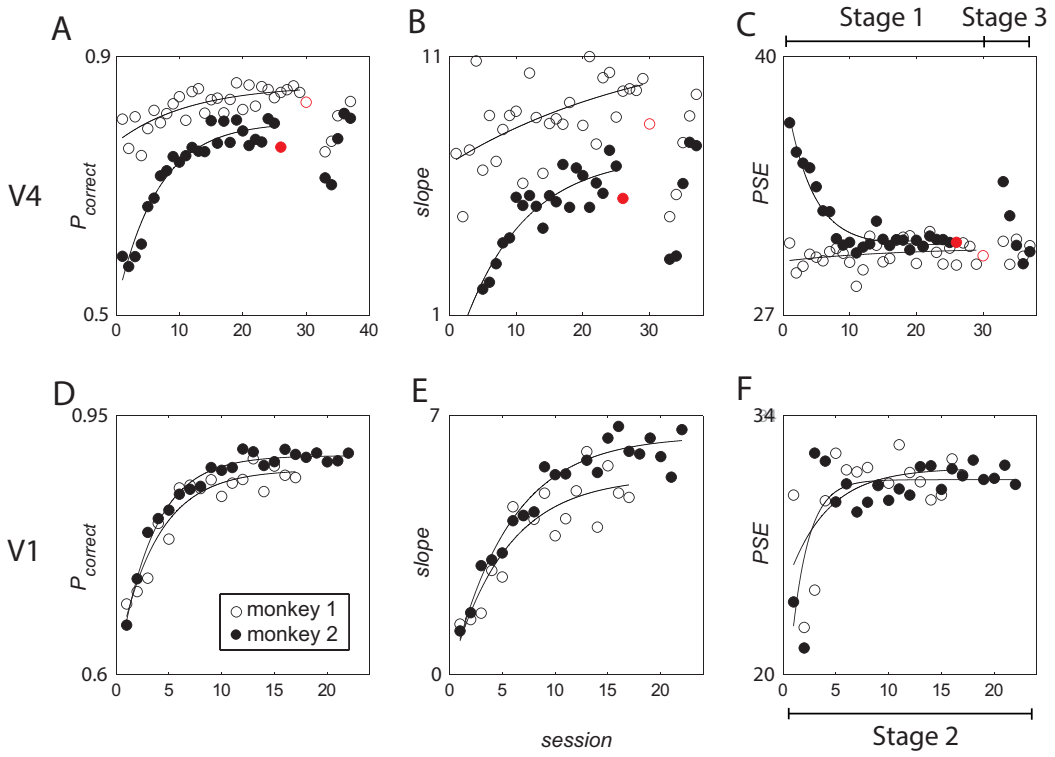


Figure 8. Performance in the contrast discrimination task over the course of training. A, B & C: V4 location (Stage 1, followed by five data points from Stage 3); D, E & F: V1 location (Stage 2). A & D: proportion of correct responses ($P_{correct}$); B & E: slope of the psychometric function (corresponding to the derivative at 30% contrast); C & F: PSE. Unfilled dots: monkey 1; filled dots: monkey 2. Black markers: vertically-oriented stimuli; red markers: horizontally-oriented stimuli. Black lines depict the best-fit exponential curves. Note that the test contrasts used in Stages 1 and 3 were identical, hence they are depicted on the same subplots.

Mean task performance, M, was compared between the first and last 30% of sessions (M_{early} and M_{late}) within each stage. For both subjects and both stimulus locations, the proportion of correct trials and the slope were significantly higher for later sessions, compared with earlier ones (monkey 1, slope at the V4 location: $t(8) = -4.68$, $q = .00184$; $P_{correct}$ at the V4 location: $t(8) = -6.34$, $q < .001$; slope at the V1 location: $t(6) = -4.67$, $q < .001$; $P_{correct}$ at the V1 location: $t(6) = -7.78$, $q < .001$; monkey 2, slope at the V4 location: $t(6) = -13.3$, $q < .001$; $P_{correct}$ at the V4 location: $t(6) = -7.78$, $q < .001$; slope at the V1 location: $t(5) = -7.45$, $q < .001$; $P_{correct}$ at the V1 location: $t(5) = -4.20$, $q = .00163$, $\alpha = .05/12\times9 = .0375$; FDR corrected, unpaired two-sample t-test).

In monkey 1, the PSE did not change with training (V4 location: $t(8) = -0.96$, $q = .377$; V1 location: $t(6) = 5.32$, $q = .162$). This was likely due to a ceiling effect, as the

PSE had shifted rapidly towards 30% within the first few training sessions, leaving little room for subsequent improvement. This trend was also observed in monkey 2, for training undertaken with stimuli at V1 ($t(5) = -1.44$, $q = .154$). When stimuli were presented at the V4 location for monkey 2, the PSE was relatively high ($M_{early} = 34.1\%$) during early sessions, and it shifted significantly towards 30% over the course of training, reaching a mean value of 30.7% during late sessions ($t(6) = 5.19$, $q < .001$, unpaired two-sample *t*-test).

1.4.1.4 Psychometric thresholds for conditions with higher or lower test contrasts

The curve fitting allowed us to examine the effects of two distinct types of learning on performance, in which the parameter λ represents the associational/ attention-based component of learning (also sometimes termed the 'finger error' for experiments in which human subjects accidentally press the unintended button when indicating their response, which in our case could be termed the 'saccade direction error'), while changes in the slope and threshold represent genuine perceptual learning.

Changes in the value of λ with training can be seen in Figure 7, by examining changes in $P_{reporthigher}$ for the conditions with the highest (dark brown markers) and lowest (dark purple markers) test contrasts, respectively. When stimuli were presented at the V4 location for either monkey, the value of λ was large during early training sessions, and the number of erroneous responses decreased over the course of training, eventually reaching values of around zero (Spearman's rank correlation, monkey 1, C_L condition: $r(27) = -.582$, $q < .001$; C_H condition: $r(27) = .476$, $q = .0091$; monkey 2, C_L condition: $r(23) = -.755$, $q < .001$; C_H condition: $r(23) = .439$, $q = .0283$). At the V1 location, the value of λ tended to already be very small at the start of training, thus it only changed significantly for 1/4 comparisons (monkey 1, C_L condition: $r(17) = -.615$, $q = .0087$; C_H condition: $r(17) = .307$, $q = .230$; monkey 2, C_L condition: $r(5) = -.600$, $q = .350$; C_H condition: $r(5) = .700$, $q = .233$, FDR correction, $\alpha = .05/8 \times 5 = .0313$).

Psychometric thresholds (T_L and T_H for the C_L and C_H test contrast conditions, respectively) are shown in Figure 9. On some occasions (particularly during early training sessions with monkey 2 when stimuli were at the V4 location), performance levels did not reach 81.6%, thus a proper psychometric threshold could not be

calculated for those sessions. In these instances, threshold levels were assigned the highest possible value (T_L = 30% for C_L conditions; T_H = 100 – 30 = 70% for C_H conditions), and data points for these sessions are indicated by an unfilled circle.

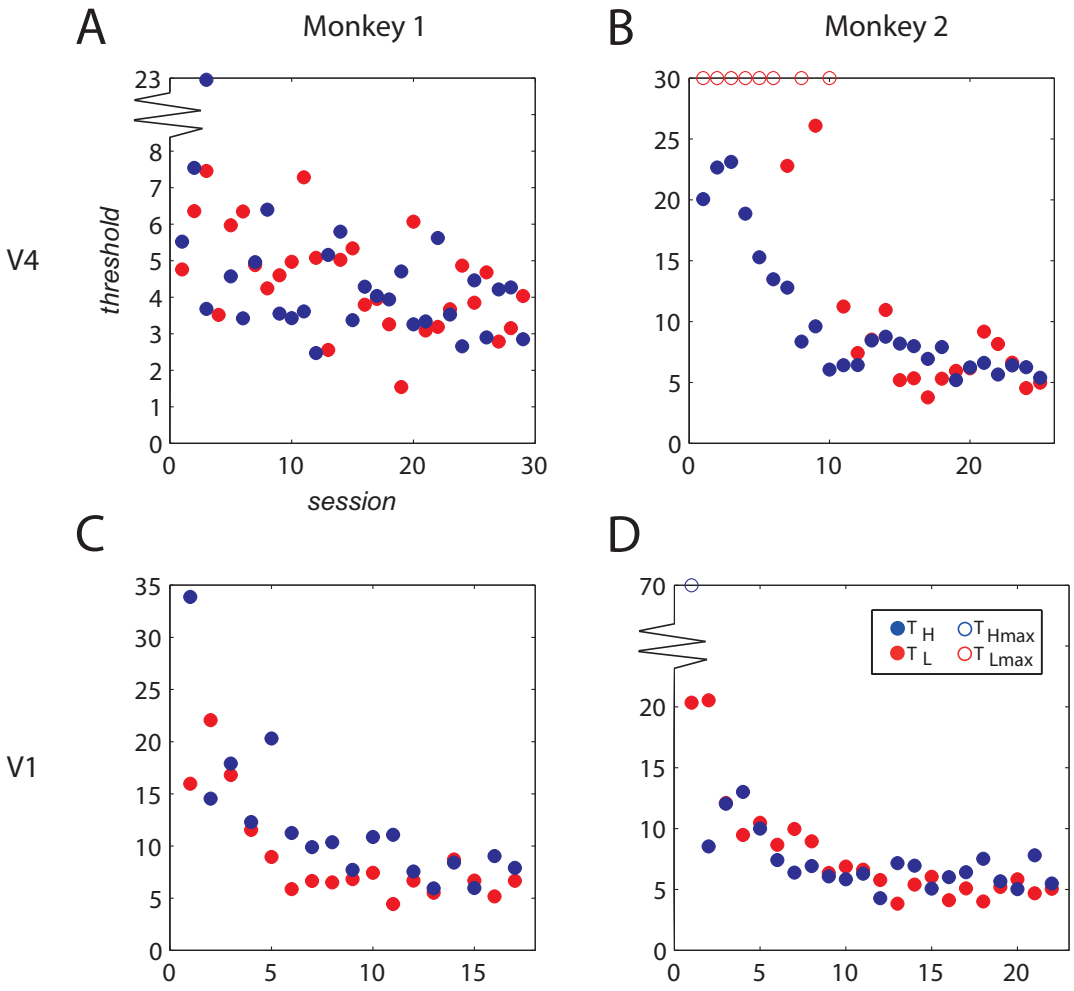


Figure 9. Psychometric thresholds, T_L and T_H, as a function of training session. A & B: V4; C & D: V1. A & C: monkey 1; B & D: monkey 2. Red markers: C_L conditions (the test contrast was lower than that of the sample); blue markers: C_H conditions (the test contrast was higher than that of the sample). Unfilled markers represent sessions in which the psychometric threshold at 81.6% could not be obtained and the threshold was thus assigned the maximum value possible (C_L conditions: T_L = 30%; C_H conditions: T_H = 70%). Significant decreases in T_L and T_H were observed in 6/8 cases (results from a Spearman's rank correlation analysis are presented in Table 3).

A Spearman's rank correlation analysis was carried out between threshold and session number, to identify changes in threshold over time. Significant decreases in

threshold were observed in monkey 1 at the V1 location and in monkey 2 at both locations (Table 2).

Statistic	*df*	*r*	*q*	*df*	*r*	*q*
	Monkey 1			Monkey 2		
	V4					
C_L	27	-.371	.0474	23	-.884	< .001*
C_H	27	-.194	.313	23	-.852	< .001*
	V1					
C_L	15	-.748	< .001*	20	-.871	< .001*
C_H	15	-.858	< .001*	20	-.623	.00195*

* $q < \alpha$

Table 2. Changes in psychometric threshold over the course of training were assessed using a Spearman's rank correlation analysis (FDR correction for α-levels: α = .05 × 6/8 = .0375).

To investigate whether condition-dependent threshold differences might be affected by the stage of training, a four-way repeated measures ANOVA was performed with condition type (T_L or T_H) as the within-session variable, and the three factors of training phase (first or second half of training sessions); subject (monkey 1 or 2); and area of stimulus presentation (V4 or V1), as the between-sessions variables. While no significant main effect of condition type was observed ($F(1,85) = 0.001$, $p = .981$), there were significant interactions between condition type and subject ($F(1,85) = 5.054$, $p = .027$); condition type and area ($F(1,85) = 11.314$, $p = .001$); condition type, training phase and area ($F(1,85) = 6.196$, $p = .015$); and condition type, training phase, subject, and area ($F(1,85) = 3.986$, $p = .049$). No significant interactions were observed between condition type and training phase ($F(1,85) = 1.339$, $p = .250$) or between condition type, subject, and area ($F(1,85) = 2.496$, $p = .118$). Upon closer examination, threshold values were significantly higher for low than for high test contrast conditions in monkey 2, when stimuli were presented at the V4 location, during the first half of training sessions (T_L: $M = 24.3$, $SE = 1.2$, 95% CI = [21.9, 26.7]; T_H: $M = 13.2$, $SE = 2.1$, 95% CI = [9.0, 17.4]). Other than this, no consistent difference between T_L and T_H was found.

1.4.1.5 Perceptual learning within individual sessions

Learning occurred across multiple sessions; could changes be detected within shorter periods of time, such as that spanned by an individual session? To investigate

this, we examined the first and last 30% of trials in a given session (termed 'beginning' and 'end' trials, respectively). The proportion of correct trials, the slope of the psychometric function, and the PSE were calculated separately for these two groups of trials. In both subjects, the proportion of correct trials was significantly higher for the last 30% than for the first 30% of trials, for training undertaken at the V1 location. (Table 3, paired *t*-test, FDR correction for α-levels, proportion correct: α =.05 × 2/4 = .025; slope: α =.05 × 1/4 = .0125; PSE: α =.05 × 1/4 = .0125).

Monkey	Location	$P_{correct}$			Slope			PSE		
		t	*df*	*q*	*t*	*df*	*q*	*t*	*df*	*q*
1	V4	-1.57	28	.129	-0.01	28	.993	1.42	28	.165
	V1	-4.52	16	< .001*	-2.68	16	.0166	-1.33	16	.201
2	V4	1.53	24	.14	1.04	24	.308	-2.32	24	.0295
	V1	-4.2	21	< .001*	-2.13	21	.0452	0.395	21	.697

* $q < \alpha$.

Table 3. Differences in performance within individual sessions. For both subjects, when performance was compared between the first and last 30% of trials, the proportion of correct responses was significantly higher towards the later part of each session, for stimuli at the V1 location.

The improvements in performance seen within sessions, for training at the V1 location, might have been due to a trade-off between speed and accuracy- the animals might have made faster responses at the beginning of each session out of impatience to receive their reward, and then slowed down as they grew satiated. To test this, we compared subjects' reaction times (RTs) between the first and last 30% of trials in each session ($RT_{beginning30}$ and RT_{end30}, respectively), for each of the training locations. When stimuli were placed in the V1 location, RTs did not differ significantly between the beginning and end of each session, for either subject (monkey 1: $t(16) = -0.0112$, $p = .991$; monkey 2: $t(21) = 1.21$, $p = .242$, paired *t*-test). Thus, the within-session improvements in performance that were observed when stimuli were in the V1 location were not due to a speed-accuracy trade-off.

When stimuli were placed in the V4 location, RTs were significantly longer at the end of each session (compared to the beginning) for monkey 1, whereas they were significantly shorter at the end of each session, for monkey 2 (monkey 1: $t(28) = 2.03$, $p = .0414$; monkey 2: $t(24) = -6.60$, $p < .001$, paired *t*-test). Thus, the lack of improvement

observed at the V4 location over the course of individual sessions could not be attributed to a speed-accuracy trade-off either, at least in monkey 2.

1.4.1.6 Reaction times

For each session, mean RTs were calculated separately for correct and incorrect trials, across all 14 test contrast conditions. RTs decreased significantly with training in monkey 1, at both the V4 and the V1 locations, for correct as well as for incorrect trials (Pearson's correlation coefficient, V4 location, correct trials: $r(27) = -.968$, $q < .001$, incorrect trials: $r(27) = -.905$, $q < .001$; V1 location, correct trials: $r(15) = -.846$, $q < .001$, incorrect trials: $r(15) = -.796$, $q < .001$). For monkey 2, significant reductions in RT occurred during training at the V4 location for correct and incorrect trials (Pearson's correlation coefficient , correct trials: $r(23) = -.715$, $q < .001$, incorrect trials: $r(23) = -.648$, $q < .001$), as well as at the V1 location for incorrect trials ($r(15) = -.409$, $q = .0241$), while a trend (non-significant) towards a decrease in RT was seen at the V1 location for correct trials ($r(15) = -.479$, $q = .059$).

1.4.2 Control task with horizontally-oriented Gabor stimuli at the V4 location

To determine whether contrast discrimination levels remained the same if the stimulus orientation was altered, horizontal Gabor stimuli were presented during a single control session (indicated by red markers in each of the upper subplots in Figure 5).

By and large, the change from vertical to horizontal Gabors did not have much effect on the monkeys' performance during the control session (X_h), indicating that learning was not specific to stimulus orientation (see Table 4).

	Monkey 1			Monkey 2		
	Late Stage 1 sessions, range $X_{min} – X_{max}$	Horizontal Gabor session, X_h	Last vertical grating session, X_g	Late Stage 1 sessions, range $X_{min} – X_{max}$	Horizontal Gabor session, X_h	Last vertical grating session, X_g
$P_{correct}$	0.823 – 0.854	0.829	0.83	0.762 – 0.803	0.759	0.804
Slope	7.6 – 11.0	8.4	9.5	5.2 – 7.4	5.5	7.5
PSE	29.5 – 31.2	30.0	30.5	30.3 – 31.0	30.6	30.2
$RT_{correct}$	146 – 166	149	166	149 – 164	167	155
RT_{error}	153 – 179	156	196	154 – 172	174	156

Table 4. Comparison of subjects' performance during control sessions, against that seen at the end of Stage 1. $X_{min} – X_{max}$: Ranges of performance seen during late Stage 1 sessions, in which vertically-oriented Gabor stimuli were presented. X_h: Performance recorded during the single session in which horizontally-oriented Gabor stimuli were presented. X_g: Performance recorded during the last of the Stage 3 sessions, in which vertically-oriented *grating* stimuli were presented. Stimuli were located at the V4 location during each of these sessions.

1.4.3 Control task with sinusoidal grating stimuli at the V4 location

Stage 3 consisted of five consecutive sessions in which subjects practised a CD task with vertically-oriented sinusoidal gratings at the V4 location, allowing us to estimate the extent to which subjects had relied on cues from the perceived size of the stimulus, to carry out the task. We expected the subjects' performance during the first few sessions of Stage 3 to be relatively poor as stimulus locations had just been switched from the V1 location back to the V4 location. Thus, our analysis focused on data that was obtained from the last of these five sessions.

For the most part, subjects' performance during this session (X_g) fell within the ranges of values seen during the late phase of Stage 1 (Table 4). Thus, the monkeys' ability to discriminate contrast levels was largely comparable between sessions with Gabor and sinusoidal grating stimuli, indicating that our subjects had relied primarily on contrast differences, rather than on perceived differences in stimulus size, to complete the task.

1.4.4 Control task with stimuli of different spatial frequencies at the V1 location

After extensive training on the contrast discrimination task, an additional control experiment was carried out with monkey 2 over two testing sessions, in which sinusoidal grating stimuli of two different spatial frequencies were positioned at the V1 location. The SF of the stimuli varied randomly from trial to trial. This allowed us to assess the degree to which learning on the contrast discrimination task transferred from the trained SF (4 cpd) to an untrained SF (2 cpd). Stimulus parameters and contrast levels remained otherwise identical to those used during training at the V1 location.

When the SF differed from that used during previous training sessions, performance was worse- the proportion of correct trials was lower, and the PSE lay further away from the sample contrast (first session, SF 4: $P_{correct}$ = 0.86, slope = 5.2, PSE = 25.3; SF 2: $P_{correct}$ = 0.75, slope = 2.5, PSE = 37.1; second session: SF 4: $P_{correct}$ = 0.89, slope = 6.2, PSE = 28.1; SF 2: $P_{correct}$ = 0.81, slope = 3.0, PSE = 32.4).

Thus, task performance was consistently better when the spatial frequency was the same as that used throughout previous training sessions (at 4 cpd), than when it was altered (to 2 cpd).

1.4.5 Control task with only the test stimulus- not the sample- at the V1 location

Finally, a single testing session was carried out with monkey 2, to determine how well the monkey performed in the absence of an external reference stimulus. The test stimulus was presented at the V1 location as before, while the sample was omitted. The monkey was not explicitly instructed on how to perform the task in the absence of the sample stimulus. However, assignation of correct and incorrect targets remained the same, and the monkey was thus provided with continuous feedback regarding his choices.

Performance in terms of the mean proportion of correct trials and the slope of the psychometric function was poorer in the absence of the sample stimulus, when

compared to that attained on preceding days in the presence of the sample (performance in the absence of a sample: $P_{correct}$ = 0.78, slope = 2.5).

Importantly, however, the PSE of the psychometric function was 30.9%, i.e. still very close to the sample contrast. This indicated that the subject was able to perform the task based on an internalised contrast reference of 30%.

1.4.6 Discussion of behavioural results from the CD task

Substantial improvements were observed in our subjects' psychophysical performance, including higher success rates in their behavioural responses, steepening of their psychometric functions, and shifts in the point of subjective equality towards the contrast of the sample stimulus.

Significant progress was often observed across training sessions that spanned several weeks; it also took place within the time frame of individual sessions which lasted just a few hours. When we examined performance levels for individual test contrast levels, we found (unsurprisingly) that the more difficult the discriminations required, the longer it generally took subjects to improve.

Thus, our study demonstrates that perceptual learning can occur during adulthood for contrast discrimination tasks, thereby complementing studies which have demonstrated enhancements of contrast detection abilities in cats (Hua et al., 2010) and contrast discrimination in humans with normal vision (Adini et al., 2004; Kuai et al., 2005; Sowden et al., 2002; Yu et al., 2004; J.-Y. Zhang et al., 2008).

As we used monkey subjects, rather than humans, this imposed practical constraints on training and data collection, and our task paradigm necessarily differed from those used in human studies (e.g. by Adini et al. (2004) and Yu et al. (2004)), in several respects. While the human studies used a staircase procedure to measure thresholds at a level of 79% correct performance, in order to monitor changes in contrast discrimination abilities, such a method was infeasible in our study, as the number of trials and blocks that our subjects completed depended on their intrinsic levels of motivation, and we could not control the timing of their activities as closely as could be done with human subjects.

Furthermore, as we could not explicitly instruct our subjects on how to perform the task, part of the initial improvements seen would have been due to general task learning, rather than to fine contrast learning. In order to distinguish between changes that accompanied the learning of coarse contrast discriminations as opposed to fine ones, we adopted a curve-fitting procedure that included a term, λ, which described the error incurred during easy task conditions (Law & Gold, 2008). The value of λ was allowed to vary between sessions, and thereby accommodated potential differences in the rates of acquisition of broad and narrow perceptual skills. We found that the learning of associational/attention-based aspects of the task occurred predominantly during the early stages of training, whereas the acquisition of fine contrast discrimination abilities was more gradual and prolonged. For the hardest conditions, involving contrasts differences of just 1% to 2%, extensive training yielded maximum levels of accuracy in the range of 0.6 to 0.7 in both our monkeys. The separation of learning into these distinct components provided clear evidence that improvements were not mere indications of basic task learning, but were also driven by enhancements in fine perceptual sensitivity.

In relation to our findings, several key questions emerge: Firstly, how is perceptual learning of contrast discrimination mediated in different visual areas? Were the behavioural improvements of our subjects attributable to changes in neuronal properties at the level of V1 and V2 (Bao, Yang, Rios, He, & Engel, 2010; Carmel & Carrasco, 2008; Crist et al., 2001; Furmanski et al., 2004; Ghose et al., 2002; Hua et al., 2010; W. Li et al., 2004; Schoups et al., 2001; Schwartz et al., 2002; Thiele, 2004; Yotsumoto et al., 2009; Yotsumoto et al., 2008); the frontal cortex (Kahnt et al., 2011); parts of the parietal lobe that are related to the attention network (Mukai et al., 2007); or to some intermediate region in the visual and cognitive processing hierarchy such as V4 (Mukai et al., 2007; Raiguel, 2006; Rainer et al., 2004; Williford, 2006; Yang & Maunsell, 2004; Zivari Adab & Vogels, 2011)?

Secondly, if modulations of neuronal activity did occur, what form would they take, exactly? One might expect to find changes in the slope of the tuning curve (Raiguel, 2006), possibly with a scaling in response amplitude (a ‘response gain’) in a manner similar to that induced by attention (Schoups et al., 2001; Williford, 2006; Yang & Maunsell, 2004). Alternatively, one might observe a shift in the location of the

midpoint of the curve towards a contrast value around which a high degree of sensitivity is required (a 'contrast gain' change, as has also been found with attention) (Martínez-Trujillo & Treue, 2002; Reynolds, 2000). It is equally possible that the 'readout' of activity levels from neurons with distinct tuning preferences is altered, depending on the stimulus and task (Berens et al., 2012; Pooresmaeili, Poort, Thiele, & Roelfsema, 2010). Training might additionally be accompanied by alterations in firing rate variability (Raiguel, 2006; Schoups et al., 2001).

A close examination of neuronal mechanisms that underlie the process of contrast discrimination learning is described in the next section.

1.5 Neuronal methods

1.5.1 Data processing

After the initial stage of data acquisition, a lengthy process of spike thresholding and artifact removal was carried out, in order to obtain spikes for further analysis. As the details of this procedure are long and rather involved, the full description is presented separately, in Appendix A: *Artifact removal from neuronal data*, page 213. However, at this juncture, it is necessary to examine how the particular methods used in the selection of spike thresholds have defined the scope of inferences that can be drawn from our data.

In our study, spike thresholds were systematically selected such that the levels of spontaneous activity obtained from the resulting spikes (that is, the activity prior to sample onset) would be uniform across sessions, for a given channel (refer to the section, '*Automated threshold setting to obtain uniform spontaneous activity levels* across sessions,' page 218). As spontaneous activity levels were deliberately kept uniform across training days, we were not able to determine whether spontaneous activity levels changed during training. What this method did allow, however, was the rigorous comparison of levels of stimulus-evoked activity across the training period, *relative* to spontaneous levels. Consequently, should changes in the shape of the contrast response function emerge over the course of training, we would not be able to distinguish whether this was best described by a response gain or an additive model (though it might still be possible to discern the effects of a contrast gain).

We additionally note that the majority of analyses and results presented in this thesis were performed using a different version of the data, in which the envelope of the rectified MUA signal was calculated, without any thresholding or standardisation of spontaneous activity levels across sessions, and crucially, that this parallel analysis yielded similar results to that based on the analysis of spiking activity (Sanayei, 2013).

1.5.2 Data analysis

1.5.2.1 Contrast response functions

Levels of spiking activity were examined for changes over the course of training. Contrast-dependent firing rates during the 512-ms test presentation period were calculated for each channel, and a contrast response function (CRF) was generated by plotting spiking activity against contrast. A Naka-Rushton function was fit to the data using the method of least-squares according to the formula

$$R = R_{max} \frac{C^n}{C^n + {C_{50}}^n} + M \qquad \text{… (Equation 4)}$$

where R refers to the observed firing rate in spikes per second; R_{max} is the maximum response level; the C_{50} is the contrast at which the response elicited was 50% of the maximum; n controls the slope of the curve; and M is the level of spontaneous activity (Albrecht & Hamilton, 1982; Sclar, Maunsell, & Lennie, 1990). To identify changes in the properties of the CRF, four parameters (the slope of the function at 30% contrast, the C_{50}, and the minimum and maximum responses) were calculated for each session and a correlation was calculated between the parameter values and session number. The slope at 30% contrast was calculated as:

$$slope = \frac{dR}{dC}\left[R_{max} \frac{C^n}{C^n + {C_{50}}^n} + M\right]$$

$$= nR_{max} \frac{C^{n-1} - C^{2n-1}}{(C^n + {C_{50}}^n)^2} \text{ … (Equation 5).}$$

This process was carried out for data from individual channels, as well as for data that were combined across channels. For the latter, the mean level of activity was calculated across channels to obtain a population firing rate for each test contrast condition and each session. CRF parameters were monitored in tandem with those from the psychometric function, in order to identify correlations between psychophysical and neuronal metrics. Based on the observations from previous studies, on the modulatory effects of attention on the shape of the CRF (as described in the section, ‘Effects of attention on contrast response functions of visually-responsive neurons,’ on page 12), hypothetical learning-induced changes might include shifts in the C_{50} towards the

sample contrast; increases in the range of the CRF; and steepening in the slope of the CRF (illustrated in Figure 10).

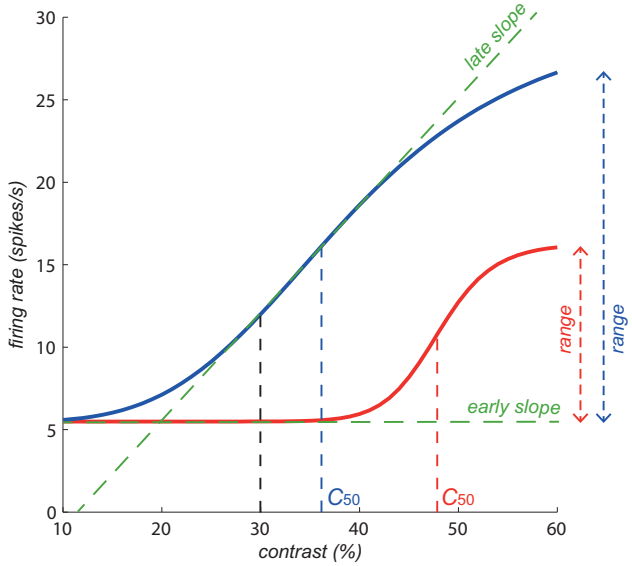


Figure 10. Illustration of hypothesised changes in the CRF with training, from early (red) to late (blue) sessions: a steepening of the slope of the CRF at 30%; an increase in the range of the CRF, and a shift in the C_{50} towards the value of 30%.

1.5.2.2 Area under the receiver operating characteristic curve (AUROC) calculation

AUROC values were calculated from spiking activity on trials where subjects made a saccade to a target (regardless of whether the response was correct or not). Firing rates were measured during each of the 512-ms stimulus presentation periods. Two approaches were adopted and tested in the calculation of AUROC values; the first approach used a traditional method of calculating AUROC values, whereas the second approach made use of a novel strategy as outlined below.

The first (traditional) method of calculating AUROC values was based on a technique borrowed from signal detection theory (Green and Swets, 1966). For each test contrast condition, the area under the receiver operating characteristic curve (AUROC) was calculated, yielding a set of AUROC values for each channel (Britten, Shadlen, Newsome, & Movshon, 1992). In detail, the ROC curve was generated by plotting the probability that the spike count of *test*-evoked responses would exceed a criterion spike count, against the probability that a spike count of *sample*-evoked response would

exceed the same criterion. After each iteration (which yielded a pair of probabilities), criteria were adjusted by 1 spike/s, until the criterion exceeded the largest spike count present in either of the two spike count distributions (corresponding to the counts for sample- and test-evoked responses). The family of probability pairs yielded a continuously distributed dataset, ranging from 0 to 1. The area under the curve corresponded to the performance of an ideal observer who was distinguishing between neuronal responses that were elicited by the sample and test stimuli.

The higher the spiking rate elicited by the test (compared to that elicited by the sample), the higher the value of the AUROC. One would expect that for an excitatory neuron, the higher the test contrast presented, the greater the AUROC value obtained, while the opposite would occur for an inhibitory neuron. To calculate AUROC values for the population response, activity was pooled across channels, and the mean levels of activity across channels were then used to generate AUROC values.

This approach involved the comparison of two distributions of activity levels, in which the degree of overlap in the distributions was quantified by the AUROC value. However, this resulted in some loss of information as it ignored potential within-trial correlations in activity. For example, consider a test contrast condition where the test always elicits a slightly higher response than the sample. Under these conditions, an ideal observer would be able to deduce which response was elicited by the test, and which was elicited by the sample, by comparing the two activity levels on every single trial. With the traditional method of calculating AUROC values, this trial-wise information is lost, and if between-trial fluctuations in activity are large relative to trial-wise differences in responses to the two stimuli, then the traditional method of AUROC calculation would yield lower AUROC values.

Thus, the basis of our second approach was to retain and exploit our knowledge of which trials yielded which pair of stimulus-evoked responses. For each trial, we asked, ‘Is the spike rate during presentation of the test stimulus higher than that during presentation of the sample?’ The fraction of trials during which the answer was ‘Yes’ corresponded to the performance of an ideal observer, giving what we termed the ‘PROBMAT’ value (‘PROBability MATching of within-trial activity’), which was calculated as

$$P_{Rt>Rs} = \frac{n_{Rt>Rs}}{n_{Rt>Rs} + n_{Rs>Rt}} \quad \ldots \text{(Equation 6)},$$

where R_t is the test-evoked response; and R_s is the sample-evoked response.

The PROBMAT values ranged from 0 to 1 and were analogous to AUROC values, except that they took within-trial correlations into account and were therefore potentially superior to AUROC calculations in determining signal separability (the distinguishing features of the two approaches are summarised in Figure 11). As with AUROC values, they could be used in the generation of neurometric functions and thresholds (described in detail in the next section).

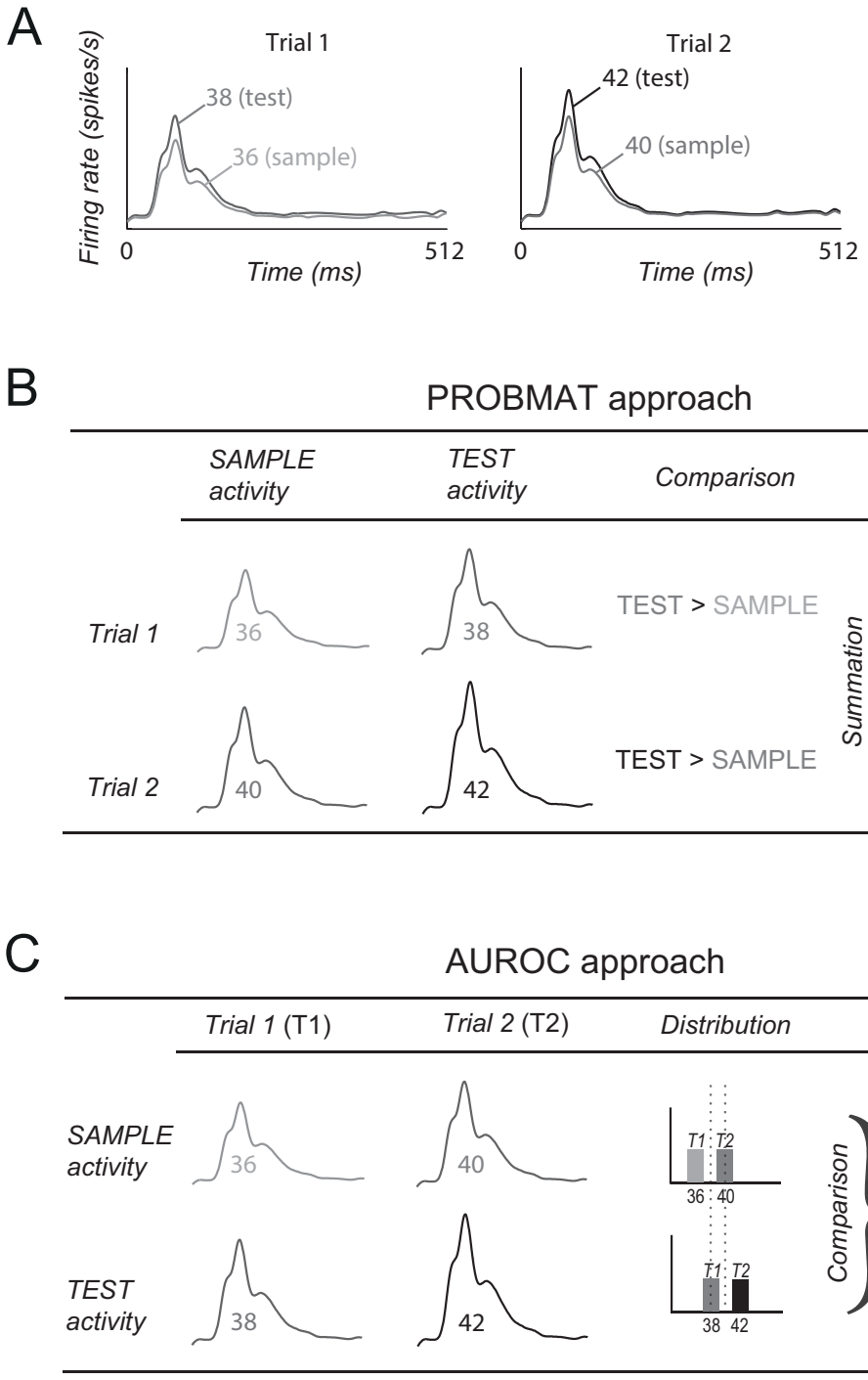


Figure 11. Illustration of the distinguishing features of the methods used to calculate the AUROC and PROBMAT measures of spiking discriminability, using data from two example trials. In this example, stimulus-evoked activity is represented by PSTHs of firing rate versus time, aligned to stimulus onset (A). During trial 1, test-evoked activity is higher than sample-evoked activity (38 > 36). During trial 2, test-evoked activity is also higher than sample-evoked activity (42 > 40). However, in trial 2, overall firing rates are systematically higher than those elicited in trial 1, by 4 spikes/s. This offset in

inter-trial firing rates may arise from factors such as ongoing fluctuations in spontaneous activity levels. In the PROBMAT approach (B), stimulus-evoked activity is compared on a trial-by-trial basis, and this process remains unaffected by trial-to-trial fluctuations as long as the relationship between test- and sample-evoked activity remains unchanged. The *fraction of trials* for which the trial-wise comparison yields 'test-higher' then yields the PROBMAT value. In the AUROC approach (C), firing rates are pooled across trials, forming separate distributions for the two stimuli. The *degree of separation* between these two distributions is then compared, producing an AUROC value. In this example, due to trial-to-trial variations in activity, the firing rate elicited by the test on trial 1 is lower than that elicited by the sample on trial 2, causing an overlap in the two distributions of activity, and impairing the performance of a decoder/ ideal observer.

1.5.2.3 Weibull curve fitting of AUROC and PROBMAT values

The data for each contrast condition were fit with a four-parameter Weibull function using maximum likelihood estimation (MLE), according to the formula

$$y = 1 - \delta - (\gamma e^{-\left(\frac{x}{\alpha}\right)^{\beta}}) \qquad \text{… (Equation 7)}$$

where y is the AUROC or PROBMAT value; x is the contrast of the test stimulus; α is the threshold; β is the slope; γ is the range; and $1 - \delta$ is the maximum AUROC or PROBMAT value reached by the neurometric function.

1.5.2.4 Monitoring the neurometric function during training

Changes in the neurometric function over the course of training were monitored using six parameters: 1. The slope of the tangent at 30% contrast; 2. The point of neurometric equality (PNE); 3 and 4. The minimum and maximum values; 5 and 6. The threshold values, T_L, and T_H, at which neurometric performance was 82% correct. As with the CRF and psychometric function parameters, AUROC and PROBMAT parameters were monitored over the course of training, for possible shifts in the PNE, steepening of the slope at 30%, and increases in the range (illustrated in Figure 12).

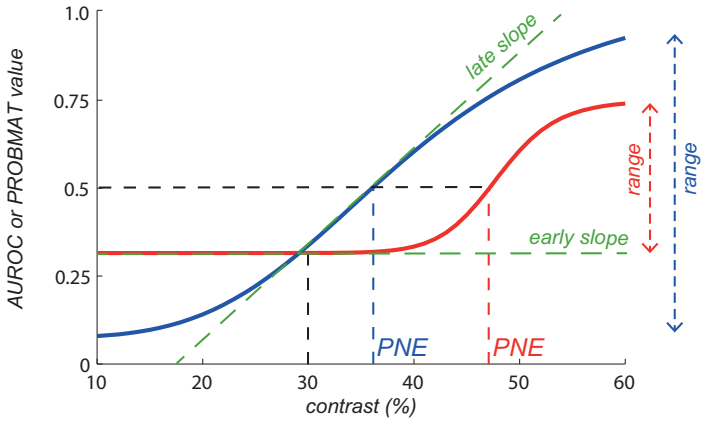


Figure 12. Illustration of hypothesised changes in the AUROC and PROBMAT functions with training, from early (red) to late (blue) sessions: a steepening of the slope of the function at 30%; an increase in the range, and a shift in the PNE towards the value of 30%.

1.5.2.5 Neurometric thresholds

Neurometric thresholds were monitored for improvements over the course of training. Based on results from a control task (previously described in the section '*Control task with only the test stimulus- not the sample- at the V1 location*,' page 37), we found that as training progressed, monkeys were likely to have performed the task by carrying out a comparison between the test contrast and an internally stored reference contrast (held in long-term memory), rather than by heeding the actual contrast of the physically-presented sample. Furthermore, results from an analysis of the effects of sensory adaptation (see the section, ' *Response adaptation prior to stimulus* onset,' page 96) implied that subjects had performed this calibration based on levels of on-going activity (i.e. just prior to test onset). Thus, for the determination of threshold levels, PROBMAT values were calculated based on a comparison of activity between pre-test and test presentation periods, rather than between sample and test presentation periods. Neurometric data were fit with Weibull functions, and the threshold was defined as the test contrast at which performance would theoretically be at 18% and 82% correct. This analysis was carried out on the neurometric 'performance' obtained from data that was pooled across the population of channels.

1.5.2.6 Response adaptation during stimulus presentation

The phenomenon of sensory adaptation typically occurs when two stimuli are presented in close succession- the response elicited by the second stimulus is often reduced due to the presentation of the first stimulus, compared to what it would have been if the second stimulus had been presented on its own.

Effects of visual response adaptation, if present, were expected to be most easily detectable during conditions where the test and sample contrasts were very similar. To determine whether adaptation had occurred, firing rates from each channel were compared between periods of sample and test stimulus presentation, for each of the conditions where the test contrast was just above 30%. Thus, the number of conditions examined varied depending on the recording site and the sample contrast (V4 location: test contrasts of 31, 32 and 33%; V1: test contrast of 32%).

A *t*-test was performed for each channel to find out whether the means of the two distributions of activity (in which each session contributed one data point) differed significantly. In addition, a *t*-test was performed for the population data, which combined activity across channels as well as across sessions.

Next, the mean population activity was calculated by taking the average across channels for each condition of interest, and an adaptation index (*AI*) was calculated according to the formula

$$AI = \frac{test\ activity - sample\ activity}{sample\ activity} \quad \ldots \text{(Equation 8)},$$

providing a measure of the difference in firing rates elicited by sample and test stimuli. To examine whether learning had an effect on adaptation, a correlation analysis was performed between *AI* and session number. An increase in *AI* values with training would imply less response adaption as learning progressed, whereas a decrease in *AI* values would indicate greater response adaption.

1.5.2.7 Response adaptation analysis prior to stimulus onset

Another form of response adaptation may have affected levels of post-stimulus activity following sample offset. Levels of spiking activity were thus compared between

the two pre-stimulus periods (the 256-ms periods before sample and test onset), for each of the channels. As the test stimulus was only presented after these pre-stimulus periods, activity levels during the periods in question were not dependent on test contrast, thus responses were pooled across conditions and compared between pre-test and pre-sample periods.

Differences in activity were assessed for individual channels. In addition, to identify differences in population spontaneous activity between the two periods, paired *t*-tests were carried out on firing rates that were combined across all channels, sessions, conditions, and trials.

Next, to determine whether changes in adaptation strength occurred with training, firing rates were pooled across channels for each trial, and trials were combined across conditions, to generate a population PROBMAT value for each session. A Spearman's rank correlation analysis was calculated between PROBMAT and session number, to assess whether differences in pre-stimulus firing rates (if present) changed with time.

1.5.2.8 Test-test discriminability

In addition to changes in discriminability between sample and test stimuli, it was possible that changes might have occurred in the level of discriminability between responses elicited by test stimuli of different contrasts. This required the pooling of data across trials, thus the AUROC method was used for this portion of the analysis. Spiking activity was analysed during conditions where the test and sample contrasts were very similar, and AUROC values were calculated based on comparisons of responses between 29% and 31% test contrast conditions, in V4, and between 28% and 32% test contrast conditions, in V1. AUROC values were then plotted as a function of session number, and a Spearman's correlation analysis was performed to identify changes in AUROC with time. This was carried out using data from individual channels, as well as for population data.

1.5.2.9 Fano factor

Previously, attention has been shown to increase response reliability, as well as to modulate visually-evoked activity, in V4 (Mitchell, Sundberg, & Reynolds, 2007) and V1 (Herrero, Gieselmann, Sanayei, & Thiele, 2013) neurons. To investigate the effects of perceptual learning on the variability of neuronal responses, the Fano factor (*FF*) was used to determine whether changes in firing rate variation occurred across trials. For each session and each test contrast condition, the level of variability in test-evoked activity across trials was calculated according to the equation

$$FF = \frac{variance}{mean} \qquad \text{… (Equation 9),}$$

where the variance was measured in units of (spikes/epoch)^2, and the mean was in units of spikes/epoch. The *FF* was monitored over time, for individual channels, as well as across channels. A two-way ANOVA, with training period (first 30% versus last 30% of sessions) and test contrast condition as factors, was performed to identify significant changes in *FF*, for both individual channel and population data.

1.5.2.10 Choice probability

Thus far, analyses focused on the degree to which recorded activity reflected the contrast levels of the task stimuli. Evidence for weak modulations of activity as a function of the monkey's upcoming choice have been found in V4 (Cohen & Maunsell, 2010), raising the question of whether such effects might be present in our data, and if so, whether they changed in strength over the course of training.

As such, choice probabilities were monitored over the course of training to assess the degree to which the monkeys' neuronal activity reflected the identity of their chosen target. Levels of spiking activity during the test stimulus presentation period were categorized according to whether the subject made a saccade to the black or to the white target. CPs were calculated between each of the two groups of activity (using standard AUROC methods), for the challenging test contrast conditions (V4: conditions with test contrasts of 27, 28, 29, 31, 32 and 33%; V1: conditions with test contrasts of 22, 25, 28, 32, 35%, and 40%). For each channel, the mean CP (for a given test contrast) was calculated for early and late sessions (the first and last five days of

training, respectively). A two-factor ANOVA was performed to determine whether CPs changed significantly with training, with training period (early versus late sessions) and test contrast as factors. In addition, for each of the different test contrasts, a post-hoc one-sided *t*-test was performed to determine whether the means of the two distributions differed significantly. A one-sided test was used as we were interested solely in whether neuronal activity became *more* indicative of the monkeys' upcoming choice during the final stages of training.

1.6 Neuronal results

1.6.1 Contrast response function analysis

1.6.1.1 Changes in the CRF for individual channels

Four parameters of the contrast response function were calculated for each session and a Spearman's rank correlation analysis was performed to identify any changes in the values of these parameters with training. These parameters were selected for the following reasons:

1. The slope of the tangent to the best-fit line at a contrast level of 30% provided a measure of how well the neuronal spiking activity was able to represent subtle differences in contrast around the contrast of the sample stimulus. The steeper the slope, the better the neuronal sensitivity, in terms of absolute firing rates.
2. The C_{50} corresponded to the contrast that elicited half of the maximum response, thus allowing the detection of shifts in contrast sensitivity.
3. The minimum and maximum values of the CRF provided a measure of absolute levels of contrast-dependent activity.

Values of each parameter were plotted against time (refer to Figure 13 for examples of channels on which significant changes occurred with training). The slope of the CRF at 30% contrast provided a measure which could be compared against the slope of the psychometric function at 30% contrast; similarly, the C_{50} could be compared to the PSE of the psychometric function. If changes at the neurometric level mirrored those seen in the behavioural data, then one would expect the slopes of the CRF and the psychometric function to increase in tandem, and the C_{50} to shift towards the value of 30%, as was seen for the PSE (Figure 8).

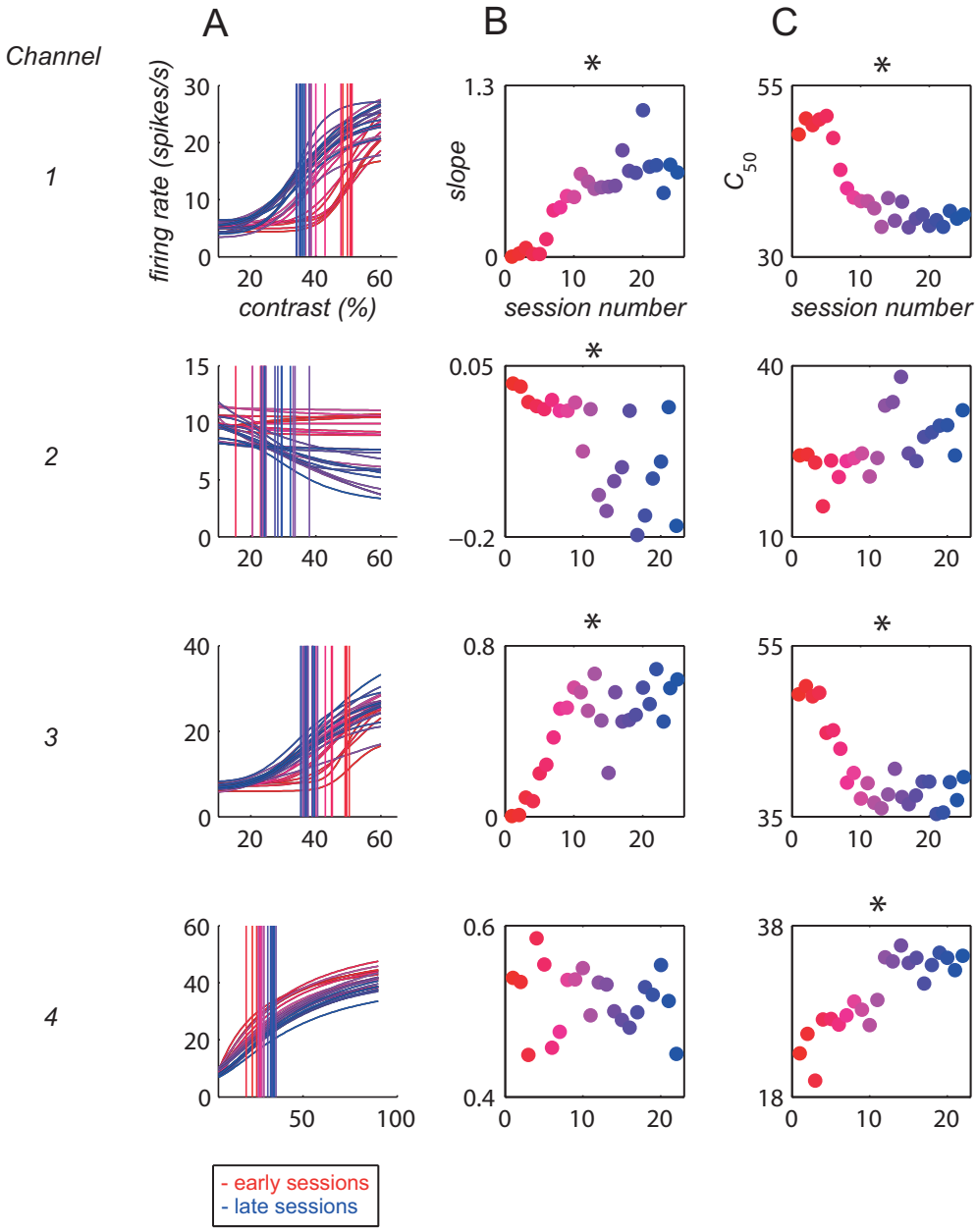


Figure 13. Changes in the CRF for four example channels (each row depicts data for one channel). Column A: Fitted curves within each subplot correspond to the CRFs obtained from multiple sessions (early sessions in red, late sessions in blue). Column B: slope of the CRF against session number; column C: C_{50} against session number. Significant changes in the slope and the C_{50} are indicated by asterisks. Increases in slope were observed in channels 1 and 3, while a decrease occurred in channel 2. The C_{50} decreased significantly towards 30% for channels 1 and 3, while it increased towards (and overshot) 30% in channel 4. Channel 1: monkey 2, V4; channel 2: monkey 1, V4; channel 3: monkey 2, V4; channel 4: monkey 2, V1.

For multiple V4 channels in both monkeys, and multiple V1 channels for monkey 2, the slope of the CRF at 30% was seen to steepen significantly (Table 5). Shifts in the C_{50} towards 30% occurred for V4 channels in both monkeys, whereas shifts occurred both towards and away from 30% for numerous V1 channels in monkey 2 (see Figure 14 for an illustration of the variety of channels that showed significant changes in the CRF). In V4, maximum firing rates increased for the majority of channels in monkey 1, and decreased for the majority of channels in monkey 2. In V1, both decreases and increases in the maximum and minimum were observed on channels in monkey 2, while little change in the minimum and maximum were seen in monkey 1.

		V4		V1	
		Monkey 1	Monkey 2	Monkey 1	Monkey 2
Slope versus session	Steeper	7	16	1	6
	Shallower	6	0	2	1
C_{50} versus session	Towards 30%	5	15	1	8
	Away from 30%	1	0	0	12
Minimum versus session	Increase	3	10	0	4
	Decrease	4	6	3	2
Maximum versus session	Increase	6	1	0	5
	Decrease	1	4	0	11

Table 5. Number of channels with significant changes for different parameters of the contrast response function, during training with sample stimuli (monkey 1, V4: $N = 29$; V1: $N = 23$; monkey 2, V4: $N = 20$; V1: $N = 25$). An FDR correction was carried out for multiple parameter comparisons.

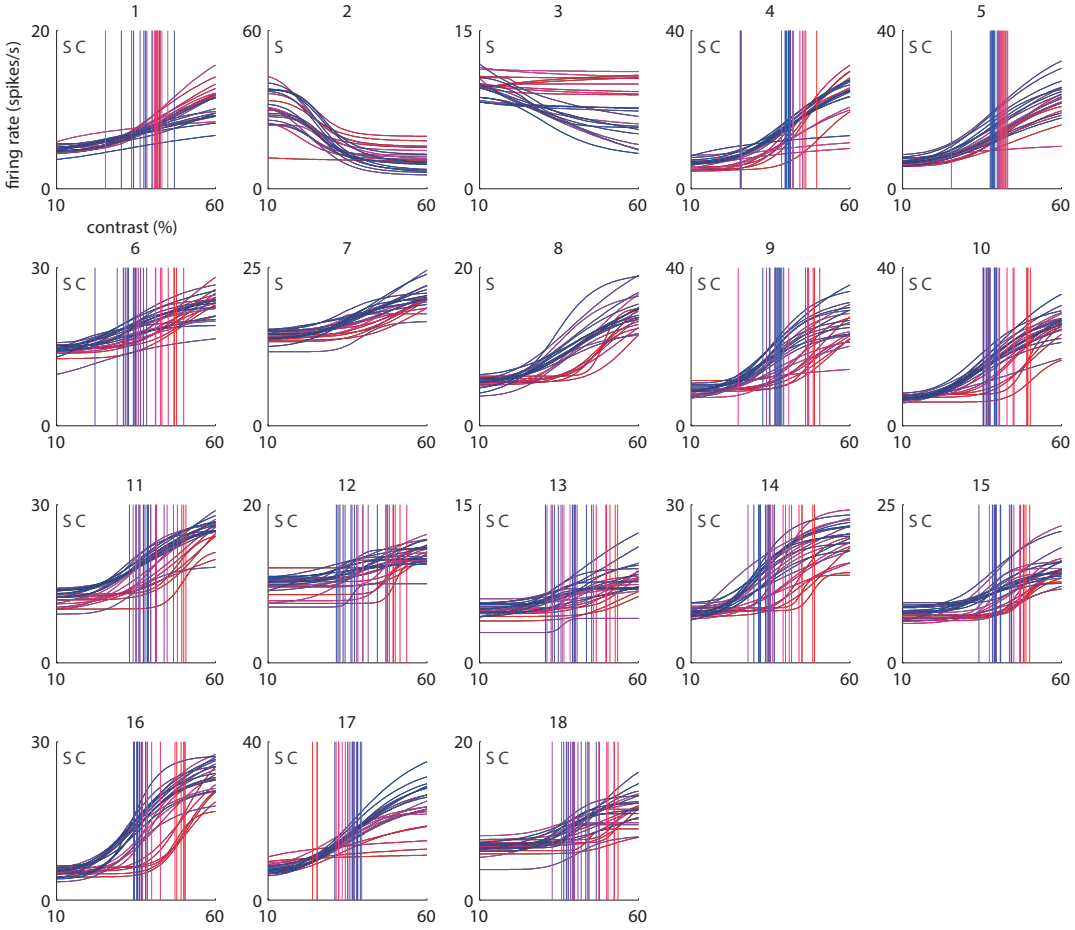


Figure 14. Changes in the CRF with training, for 18 example V4 channels. Fitted curves within each subplot correspond to the CRFs from multiple sessions (early sessions in red, late sessions in blue). The x-axis shows the contrast of the test stimulus; the y-axis shows the firing rate for a given test stimulus (spikes/sec). Increases in slope were present for each of the channels depicted (indicated by an 'S'), and many channels also showed changes in C_{50} (indicated by a 'C'). For channels with significant changes in the C_{50}, vertical lines demarcate the location of the C_{50} for each session. Across the board, shifts in the C_{50} consistently occurred in the direction of 30%.

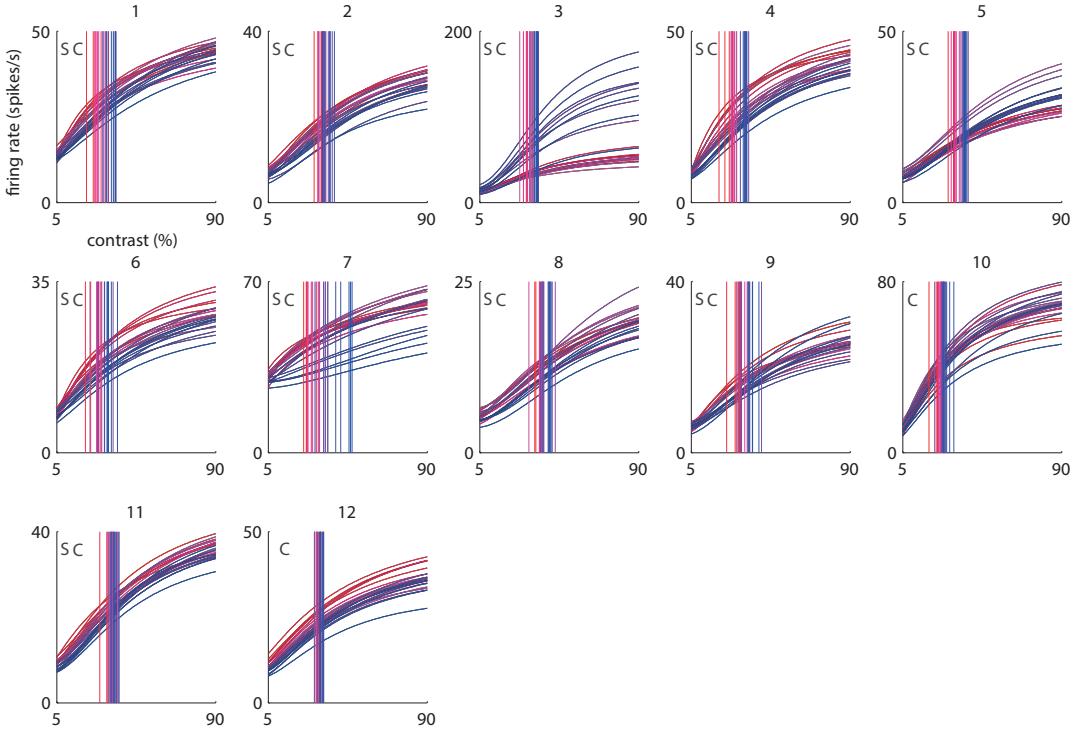


Figure 15. Changes in the CRF with training, for 12 example V1 channels. Fitted curves within each subplot correspond to the CRFs from multiple sessions (early sessions in red, late sessions in blue). The x-axis shows the contrast of the test stimulus; the y-axis shows the firing rate for a given test stimulus (spikes/sec). Increases in slope were present for most of the channels depicted (indicated by an 'S'), and all channels showed changes in C_{50} (indicated by a 'C'). For channels with significant changes in C_{50}, vertical lines demarcate the location of the C_{50} for each session. Across the board, shifts in the C_{50} were consistently towards the right, which could be in the direction of or away from 30%, depending on the channel. On some channels, e.g. channel 7, the C_{50} initially started below 30%, and then shifted towards and 'overshot' 30%.

1.6.1.2 Changes in the CRF based on the cumulative firing rate across channels

Population activity was calculated by combining spikes across channels and trials, for each test contrast condition. Mean responses were plotted against contrast and fitted with a Naka-Ruston function, to generate population CRFs for each session (Figure 16).

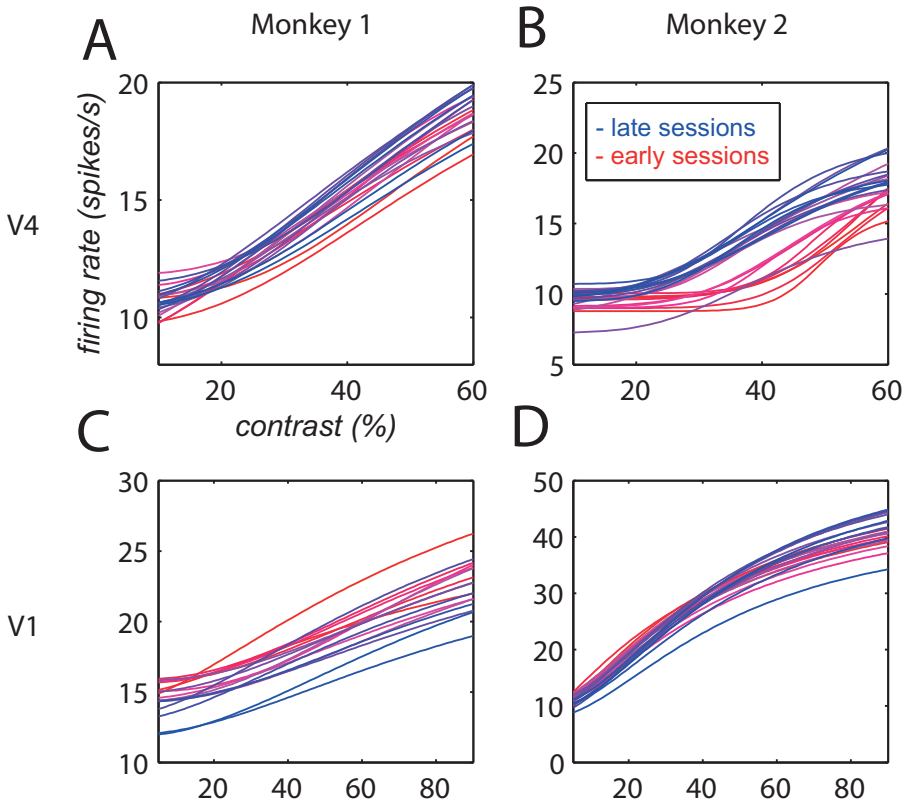


Figure 16. Population CRFs, where each fitted curve corresponds to one session (early sessions in red, late sessions in blue). A & B: V4; C & D: V1. A & C: monkey 1; B & D: monkey 2. The x-axis shows the contrast of the test stimulus; the y-axis shows the population firing rate for a given test stimulus (spikes/sec).

Parameters of the population CRF were plotted against time (Figure 17), and a Spearman's rank correlation was performed to identify training-induced changes. At both locations, results at the level of individual channels matched those seen in the population data. In V4, training was accompanied by a steepening in the slope and a shift in the C_{50} towards 30% in monkey 2, whereas no clear trend for an increase or decrease in the slope or a shift in the C_{50} was seen for the population CRF in monkey 1 (Table 6).

In V1, as with the individual channel data, changes were seen for monkey 2, but less so for monkey 1. In monkey 2, the slope increased with training, in a similar manner to that seen for the population CRF in V4. However, in this recording site, the C_{50} shifted *away* from 30%- the opposite direction from that predicted, if neurometric

performance were to match psychometric performance. In monkey 1, a significant decrease was seen in the minimum.

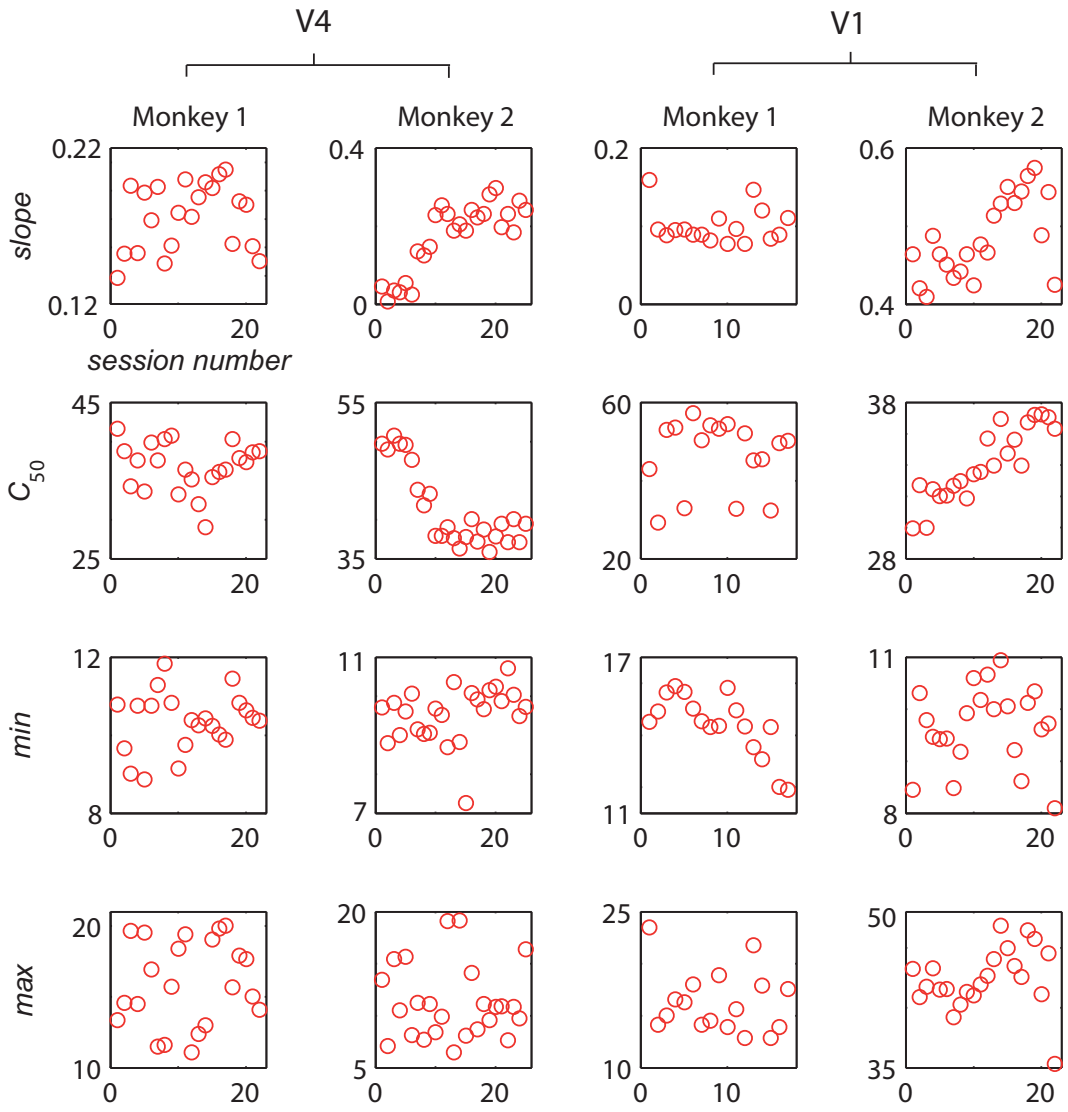


Figure 17. Parameter values of the contrast response function with time, for population activity (averaged across channels prior to fitting). First and second columns: V4; third and fourth columns: V1. First and third columns: monkey 1; second and fourth columns: monkey 2. First row: slope; second row: C_{50}; third row: minimum value; fourth row: maximum value. Significant changes were seen in the slope and the C_{50} for monkey 2 at both locations (see Table 6). After the exclusion of channels which showed stimulus-evoked suppression of activity, a non-significant trend for an increase in slope was seen for monkey 1 at the V4 location (see the section, 'Exclusion of channels with stimulus-evoked suppression,' page 60).

	Spearman's rank correlation					
	df	*r*	*p*	*df*	*r*	*p*
	V4 location					
	Monkey 1			Monkey 2		
Slope	20	.19	.386	23	.76	< .001*
C_{50}	20	-.07	.755	23	-.71	< .001 *
Min	20	.11	.621	23	.40	.0472
Max	20	.12	.582	23	-.02	.922
	V1 location					
	Monkey 1			Monkey 2		
Slope	15	-.05	.846	20	.62	.00244*
C_{50}	15	-.09	.730	20	.90	< .001*
Min	15	-.73	.00119*	20	.09	.683
Max	15	-.22	.399	20	.31	.157

* $q < \alpha$

Table 6. Changes in the contrast response function for population activity, with training. A Spearman's rank correlation was performed to assess changes in the slope at 30%, the C_{50}, and the minimum and maximum values, of the CRF. Significant improvements were seen in the slope and the C_{50} for monkey 2 at the V4 location, while deteriorations occurred for monkey 2 at the V1 location. A decrease in the minimum was seen for monkey 1 at the V1 location (FDR correction, slope: $\alpha = .05/4\times3 = .0375$; C_{50}: $\alpha = .05/4\times3 = .0375$; minimum: $\alpha = .05/4\times2 = .025$; maximum: $\alpha = .05/4\times1 = .0125$).

While improvements at the behavioural level occurred for both subjects, the changes in neurometric performance were more marked for monkey 2, particularly at the V4 location (the population CRF slope steepened in both locations, but the C_{50} only displayed a clear shift towards the sample contrast in V4 channels). An examination of CD learning rates between the two subjects showed that improvements were more gradual for monkey 2 than for monkey 1, at the V4 location, across a range of low test contrast conditions (compare the unfilled markers, corresponding to C_L conditions, between the two monkeys in Figure 7, page 29). This difference in learning rates between the subjects at the behavioural level may have accounted for the higher degree of change observed in monkey 2, at the neuronal level.

1.6.1.3 Exclusion of channels with stimulus-evoked suppression

In a minority of instances (monkey 1, V4 location: 3/29 channels; monkey 2, V4 location: 1/20 channels), neurons displayed stimulus-evoked suppression, rather than excitation, during the period of stimulus presentation. For these channels, the higher the

contrast, the lower the firing rates elicited (the opposite pattern to that seen in channels with stimulus-evoked excitation). Should training-induced modulations of the CRF have occurred as a result of learning, one would expect the contributions of excitatory neurons to differ from those of inhibitory neurons, and these opposing effects might be masked by an indiscriminate pooling of channels across the population.

Thus, the cumulative CRF was calculated, this time including only the data from channels that showed stimulus-evoked excitation. The four parameters of the CRF were calculated and plotted as a function of training session, and a Spearman's rank correlation was calculated to identify changes with training, as before.

While the results were similar to those obtained during the inclusion of all channels, the patterns observed upon the exclusion of suppressed channels supported our hypothesis: a visible trend for an increase in the slope of the CRF emerged for monkey 1, though this was not significant after an FDR correction for multiple comparisons ($r(20) = .43$, $q = .0472$, Spearman's rank correlation).

1.6.2 Non-monotonic contrast tuning functions in V4

To examine contrast-dependent responses for each channel across the whole training period, without taking effects of training into consideration, test-evoked spiking activity was pooled across all trials and sessions, to generate fourteen PSTHs per channel (one for each test contrast condition). Two V4 channels in monkey 1 (channel 14 and 55) were observed to have non-monotonic contrast tuning responses; they exhibited a preference for intermediate, rather than high, contrast levels (Figure 18).

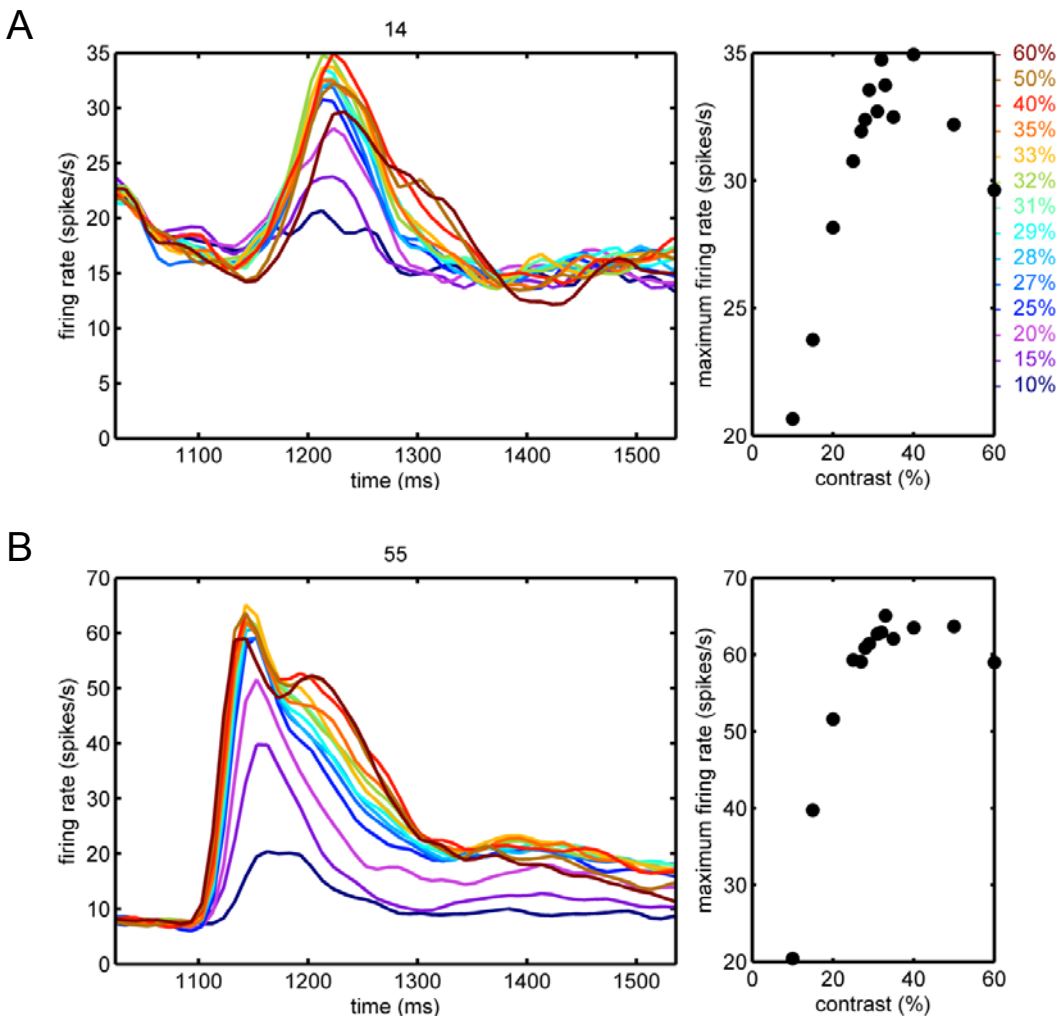


Figure 18. Left column: PSTHs showing test-evoked responses to different contrasts (colour coded by condition) for the two V4 channels in monkey 1 that exhibited non-monotonic contrast tuning responses, channel 14 (A) and channel 55 (B). Right column: Peak test-evoked firing rates as a function of contrast. The conditions that elicited the strongest responses were those with intermediate stimulus contrasts. Note that the times shown on the x-axis are measured relative to sample onset.

When the analysis was repeated with the exclusion of these two channels, the results reported in the previous section (regarding changes in the population CRF parameters with training) were not altered. No significant changes were seen in any of the four parameters for monkey 1 at the V4 location (slope: $r(20) = .257$, $q = .248$; C_{50}: $r(20) = -.390$, $q = .0724$; minimum: $r(20) = -.254$, $q = .254$; maximum: $r(20) = .251$, $q = .260$, FDR correction, $\alpha = .05/4 \times 4 = .05$).

1.6.3 AUROC/PROBMAT individual channel results

1.6.3.1 A comparison of two different methods of assessing contrast-dependent responses

Using two different methods, neurometric functions were generated by plotting AUROC and PROBMAT values against test contrast. Both measures offered a measure of the discriminability of spiking activity between sample and test stimuli, for individual recording channels. AUROC values provided an across-trial summary of stimulus-evoked discriminability, while PROBMAT values performed a similar function, but additionally took trial-wise variability into account.

We found that the PROBMAT method consistently outperformed the AUROC method at both the single-channel level and the population level. Figure 19 presents spike data, taken from the V1 location in monkey 2, for an example test contrast of 20%, in which AUROC and PROBMAT values were calculated for single channels as well as for pooled channels. In this example, responses to stimuli of 20% contrast were compared to those elicited by 30% contrast stimuli, thus AUROC and PROBMAT values were lower than 0.5 (if, on the other hand, the test had been of higher contrast than the sample, then values of above 0.5 would be expected). Channels were ordered according to their AUROC value, from those closest to 0.5 to those furthest from 0.5 (i.e. closest to 0)- the equivalent of ordering them in terms of ascending signal discriminability. For each channel, a comparison of PROBMAT (grey filled markers) and AUROC values (grey unfilled markers) revealed that the PROBMAT value lay closer to 0, whereas the AUROC value lay closer to 0.5, thus indicating that the PROBMAT approach enhanced the discriminability of spiking activity.

In addition, the data showed that the joint performance of multiple recording channels was better than that of the most informative single channel. The effect of pooling data across an increasing number of channels is demonstrated by the negative gradient in the distributions of black markers in Figure 19. The larger the number of channels included in the analysis, the lower the AUROC and PROBMAT values (i.e. to closer they lay to 0). Furthermore, the advantage conferred by the PROBMAT method was visible regardless of the number of channels included in the pool- PROBMAT

values (black filled markers) for pooled data were consistently lower than AUROC values (black unfilled markers).

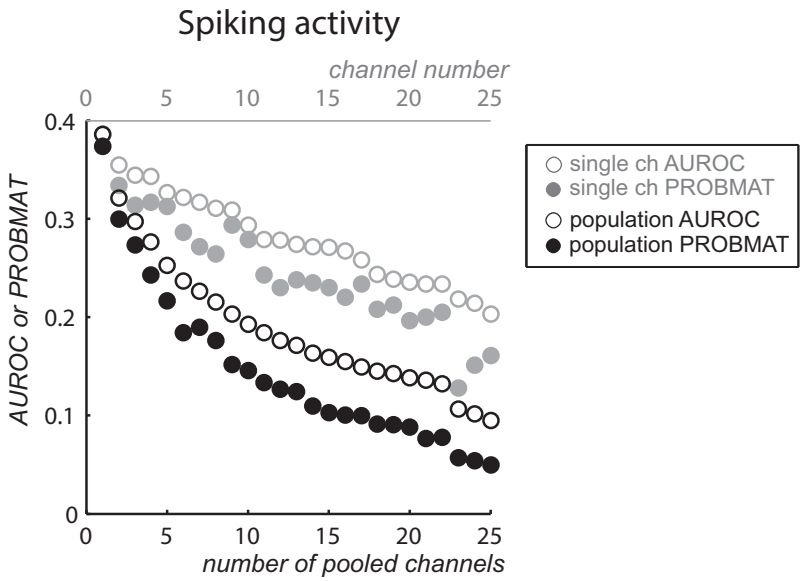


Figure 19. A comparison of the AUROC (unfilled markers) and PROBMAT (filled markers) methods of calculating ideal-observer performance for single-channel data (upper x-axis, grey) and population data (lower x-axis, black). Single-channel data are presented without any pooling across channels, while population AUROC and PROBMAT values were calculated by pooling data across increasing numbers of channels; i.e. for the population data, location 1 on the lower x-axis represents the AUROC and PROBMAT values from channel 1 only, location 2 represents data combined across channels 1 and 2, …, and location *N* represents data combined across channels 1 to *N*. These data were recorded from V1 neurons in monkey 2, for trials in which the contrast of the test stimulus was 20%. The PROBMAT method resulted in better discriminability readings, both for individual channels and for data that was pooled across multiple channels. Regardless of the approach used, decoding was enhanced by a pooling of data across channels.

In summary, the PROBMAT approach out-performed the AUROC approach, as shown by the location of unfilled dots relative to filled ones. Note that the results presented in Figure 19 were based on a pooling of trials across all the recording sessions; when a similar analysis was carried out, in which PROBMAT and AUROC values were calculated for each session and then the mean obtained across sessions, the outcome was virtually identical.

Next, the efficacy of each method was examined for all the channels, at both recording locations in the two monkeys. Figure 20 presents the PROBMAT data

obtained from each channel (small blue dots), as well as that obtained through a pooling of data across all the channels within a given recording site and subject (red and blue unfilled markers for AUROC and PROBMAT values, respectively).

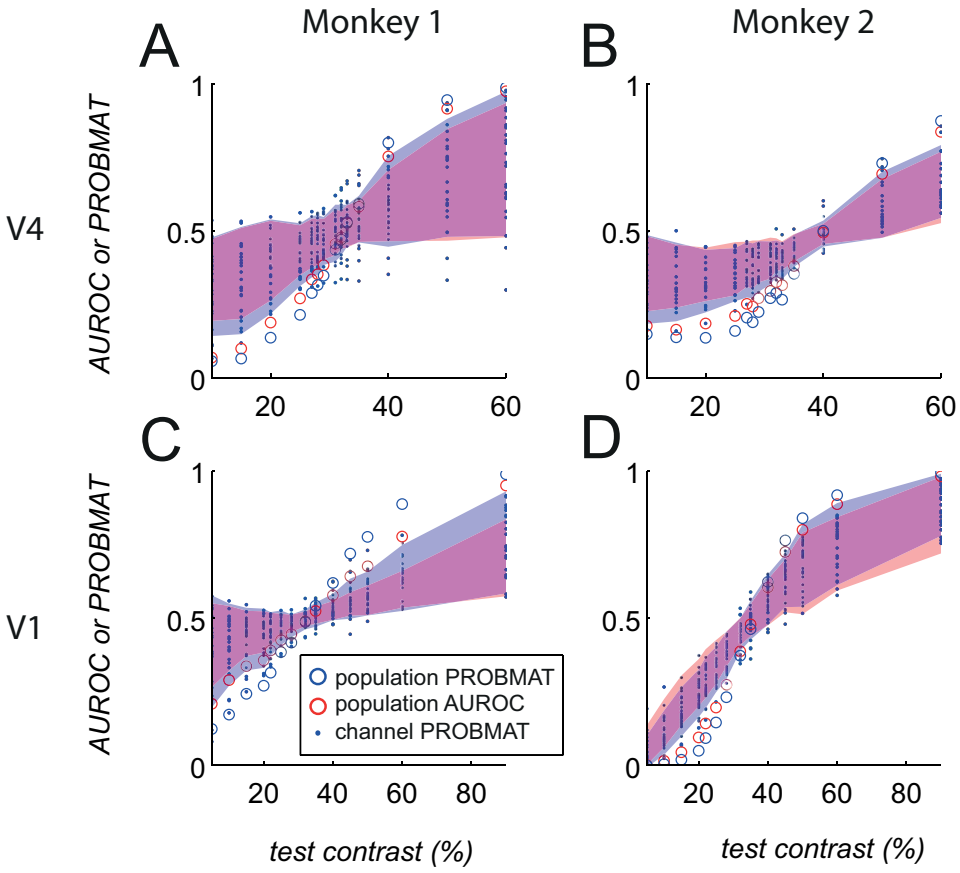


Figure 20. AUROC and PROBMAT values as a function of test stimulus contrast. A & B: V4 location; C & D: V1 location. Left column: monkey 1; right column: monkey 2. PROBMAT values for individual channel data are represented by blue dots; blue patches represent the interquartile range of PROBMAT values for individual channel data, while red patches represent the interquartile range of AUROC values for individual channel data. Population values, based on data that are pooled across all channels, are represented by blue (PROBMAT) and red (AUROC) circles. The horizontal grey line at $y = 0.5$ indicates indistinguishable neuronal responses between the two stimuli. For test contrasts below 30%, better discriminability is indicated by AUROC and PROBMAT values that lie close to zero, whereas for test contrasts over 30%, better discriminability corresponds to AUROC and PROBMAT values that lie close to one.

A visual inspection of Figure 20 revealed that for individual channel data, when PROBMAT values (blue patches) were compared with AUROC values (red patches), the interquartile range across channels tended to be shifted towards zero for test

contrasts below 30%, and towards one for test contrasts above 30%. Similarly, the population PROBMAT values occupied a wider range than did the population AUROC values, indicating that the PROBMAT approach offers better levels of discriminability. Furthermore, the population data tended to consistently outperform even the best-performing channel, indicating that the pattern observed in Figure 19 held true across the animals and recording locations- the pooling of trial-wise information across multiple channels allowed better decoding of the signal than did the monitoring of information over an extended period from even the most informative channels.

To further demonstrate the applicability of the PROBMAT approach, a Weibull function was fitted to both sets of population-derived spiking data, and the slope and range parameters were compared between the two methods. One would expect an improvement in stimulus discriminability to be accompanied by steeper slopes and wider ranges of the fitted functions. Indeed, the PROBMAT functions had significantly steeper slopes than those of the AUROC functions (paired *t*-test, $t(7) = 4.00$, $p = .0052$). Similarly, the ranges of the PROBMAT functions were consistently wider than those of the AUROC functions, although this trend was not significant (paired *t*-test, $t(7) = 2.16$, $p = .0673$).

1.6.3.2 Changes in PROBMAT values with training

The above analysis showed that the PROBMAT method yielded benefits over the AUROC method (further elaborated upon in a later section, '*A comparison of population AUROC and PROBMAT values*,' page 77). Due to the superiority of the PROBMAT method in maximising the discriminability of responses to sample and test stimuli, the rest of the analysis presented in this section was carried out using PROBMAT values.

As with the analysis performed to identify changes in the CRF over time, four parameters that described the neurometric curve (slope, PNE, minimum and maximum) were calculated and a Spearman's rank correlation was performed to identify changes in the values of each of these parameters over the course of training. These parameters were selected for the following reasons:

1. The slope of the tangent to the best-fit line at a contrast level of 30% provided a measure of how well the neuronal spiking activity was able to represent subtle differences in contrast around the contrast of the sample stimulus, and was analogous to the slope of the CRF. The steeper the slope, the better the neural discriminability, towards test and sample stimuli.
2. The PNE provided an indication of the contrast at which the AUROC value reached 0.5, i.e. the contrast at which responses to sample and test stimuli were indistinguishable. As with the C_{50} measure from the CRF, it was hypothesised that shifts in the PSE towards 30% might be accompanied by similar shifts in the PNE.
3. The minimum and maximum values of the neurometric function provided a measure of the spread of discriminability, for a given range of contrasts.

During a subset of sessions for some channels, the range spanned by the PROBMAT values did not include the value of 0.5 (i.e. the fitted neurometric curve was located entirely within either the upper or lower half of the range spanned by the y-axis), thus the PNE could not be calculated for these sessions. Channels were included in this section of the analysis if the PNE could be calculated on at least 80% of sessions, resulting in the inclusion of 15/29 V4 and 21/23 V1 channels from monkey 1, and 11/20 V4 and 25/25 V1 channels from monkey 2.

Changes in the slope of the PROBMAT curve were observed in each of the subjects, at each recording site, in a mixture of directions across the two monkeys (Table 7 and Figure 21). Significant shifts in the PNE were found on some channels at each location: for V4 channels, the PNE shifted towards 30% in all of the cases where a significant shift was observed, whereas for V1 channels, shifts usually occurred away from 30% (see Figure 21 and Figure 22 for examples of channels in which significant changes in the neurometric function occurred).

		V4		V1	
		Monkey 1	Monkey 2	Monkey 1	Monkey 2
Slope versus session	Steeper	4	9	0	4
	Shallower	3	0	2	0
PNE versus session	Towards 30%	6	5	0	2
	Away from 30%	0	0	1	5
Minimum versus session	Increase	1	0	2	2
	Decrease	2	10	0	1
Maximum versus session	Increase	3	0	1	6
	Decrease	1	5	0	1

Table 7. Number of channels with significant changes in each parameter of the neurometric function, during training on the contrast discrimination task (monkey 1, V4: $N = 15$; V1: $N = 21$; monkey 2, V4: $N = 11$; V1: $N = 25$). An FDR correction was carried out for multiple parameter comparisons.

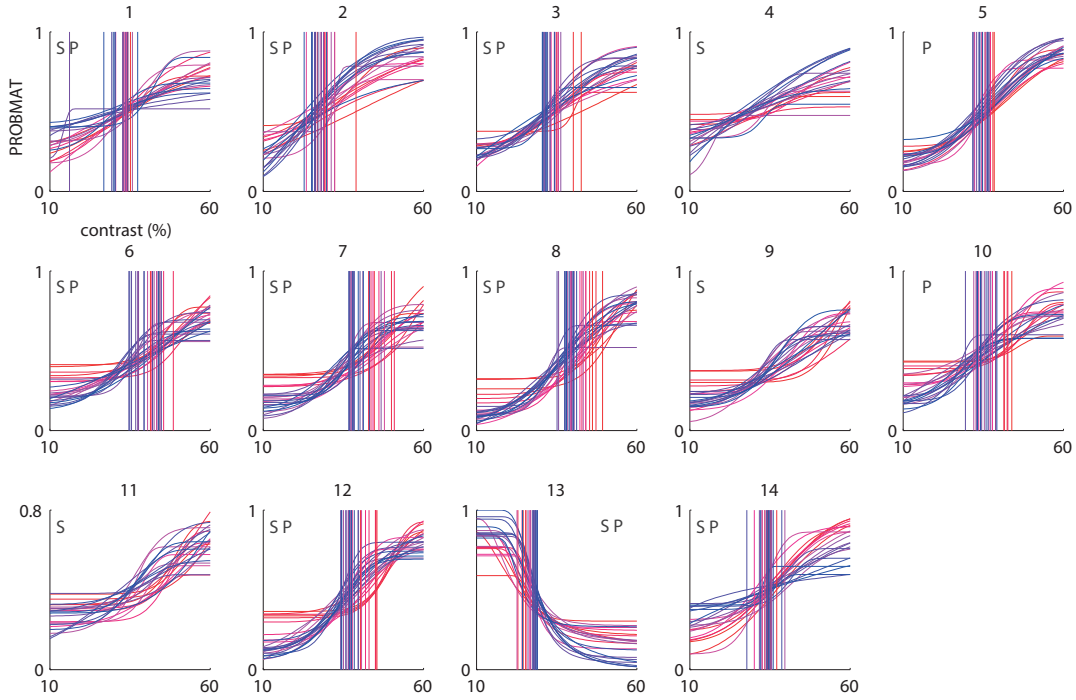


Figure 21. Neurometric functions across sessions, for example V4 channels (numbered 1 to 14) that showed significant changes in the slope (marked by an 'S') and the PNE ('P') of the PROBMAT function over the course of training. Subplots depict the fitted curves across sessions, from early (red) to late (blue). On the majority of channels, the slope increased with training, while in a minority of cases, decreases in slope were seen (subplot 14). In one case, the slope became more negative (subplot 13); this channel exhibited stimulus-evoked suppression, rather than excitation. For most of the V4 channels, the PNE started above 30%, and decreased towards 30% over the course of training. The one exception was a channel with stimulus-induced suppression (subplot 13), in which the PNE started below 30% and increased towards 30%.

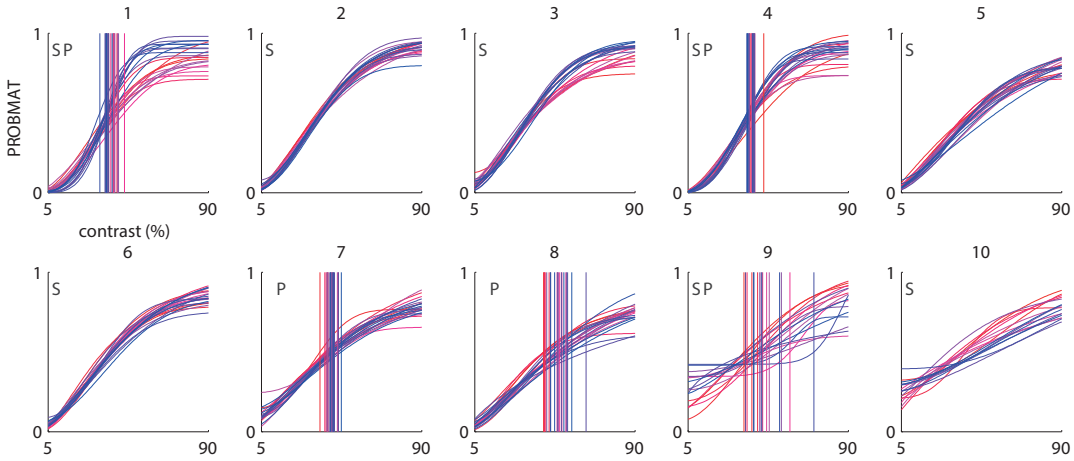


Figure 22. Neurometric functions across sessions, for example V1 channels (numbered 1 to 10) that showed significant changes in the PROBMAT function over the course of training. Conventions follow those used in Figure 21. On the majority of channels, the slope increased with training, as shown by the steepness of the blue curves relative to the red ones, while in a few cases, decreases in slope were seen (subplots 9 and 10). In the V1 channels, the PNE tended to increase *away* from the value of 30%, such as in subplot 9 (the opposite trend from that seen in V4).

We hypothesised that when the PNE shifted away from 30%, as seen in some V1 channels, this could potentially serve to broaden the dynamic range across the population of channels. To identify shifts in the PROBMAT function, a measure, C_{Half}, was calculated as the contrast at which the PROBMAT function reached half of the maximum value, $\mathrm{PROBMAT}_{\mathrm{half}}$. $\mathrm{PROBMAT}_{\mathrm{half}}$ was calculated according to the formula

$$PROBMAT_{half} = \frac{PROBMAT_{max} + PROBMAT_{min}}{2} \quad \ldots \text{(Equation 10).}$$

Sessions were divided into two groups (consisting of the first and last 30% of sessions), and the distributions of individual channel C_{Half} values were plotted for each of the groups (Figure 23).

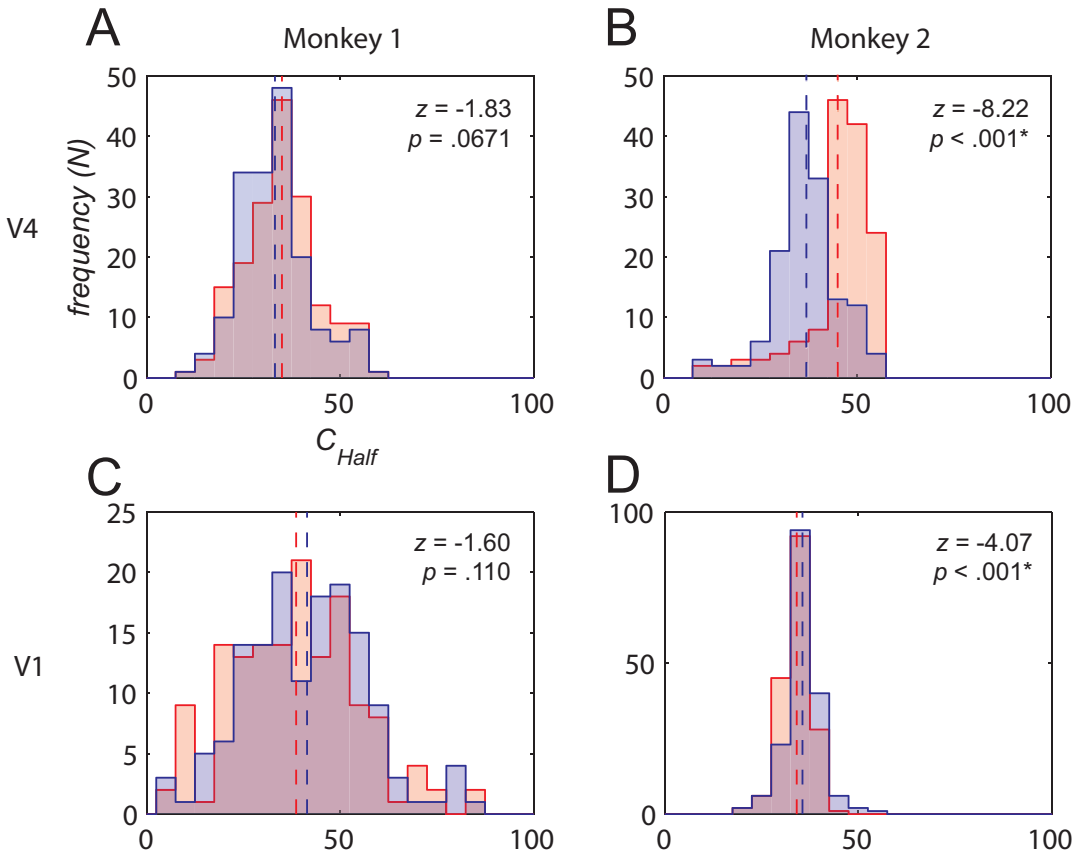


Figure 23. Distributions of C_{Half} values for individual channels, during early (red) and late (blue) sessions. A & B: V4; C & D: V1. A & C: monkey 1; B & D: monkey 2. Significant decreases in C_{Half} were observed for channels in monkey 2 at the V4 location (B), over the course of training. Vertical dotted lines indicate the means of the respective distributions.

To investigate whether the distribution of C_{Half} values across channels changed during training, a Wilcoxon signed rank test was performed to identify differences in individual channel C_{Half} values between early and late sessions. This revealed a significant difference between C_{Half} values for monkey 2, at both V4 and V1 locations. Over the course of training, C_{Half} values decreased towards the value of 30% in V4, but increased *away* from 30% in V1. This result clearly matched that seen in the individual channel PNEs.

Next, a Levene's test for equality of variance was conducted to determine whether the variances of the distributions of C_{Half} values differed between early and late sessions. A significant increase in the level of variance in the data was observed for monkey 2 in area V1 (monkey 1, V4: $F(1,346) = 0.261$, $p = .610$, V1: $F(1,228) = 0.180$,

q = .672; monkey 2: V4: $F(1,278) = 0.594$, q = .441, V1: $F(1,298) = 4.74$, q = .0303, Levene's test for equality of variance). This corresponded to a broadening in the width of the distribution, thus supporting our hypothesis that the initially counter-intuitive shifts in the PNE away from 30%, that were seen in V1 for monkey 2, may have served as a decoding strategy. A broadening of the range might have led to a reduction in the number of neurons that were highly sensitive to a given range of contrasts, which in turn might have aided decoding of the signal.

1.6.4 AUROC/PROBMAT population results

1.6.4.1 A comparison of two methods of generating population PROBMAT values

The PROBMAT method relied on a trial-by-trial comparison of test- and sample-induced activity. For the analysis of individual channel data, the calculation of PROBMAT values was straightforward. Population PROBMAT values, on the other hand, could be calculated in two ways.

The first option was to generate PROBMAT values separately for individual channels, and thereafter to calculate the mean PROBMAT values across channels. This treated the responses of individuals as separate contributions- the pooled (averaged) activity could be no better than the best channel; rather, it provided an impression of the mean response among sampled channels.

The second option was to pool activity across multiple channels, prior to calculation of PROBMAT values. This meant that information across the population was combined during each trial. Even if some neurons failed to accurately encode the stimulus contrast on some trials, the accumulation of information across the population would compensate for their performance (for a review on the robustness of population codes, see Pouget, Dayan, and Zemel (2000)). Assuming relatively low trial-wise correlations across neurons (see the section, 'PROBMAT and noise correlations,' page 88), the pooling would enhance the stimulus-encoding abilities of the population.

The latter method also appeared to offer a more biologically-plausible mechanism, as it reflected the fact that subjects had access to information across a large pool of neurons at any given time, whereas they were unlikely to depend solely on the

information from single neurons. To investigate the efficacy of the two methods, PROBMAT values were calculated using the first (*Pmean*) and second (*Pcumulative*) methods. PROBMAT values were fit with a Weibull function and the slope, PNE, minimum and maximum were plotted against session number (Figure 24).

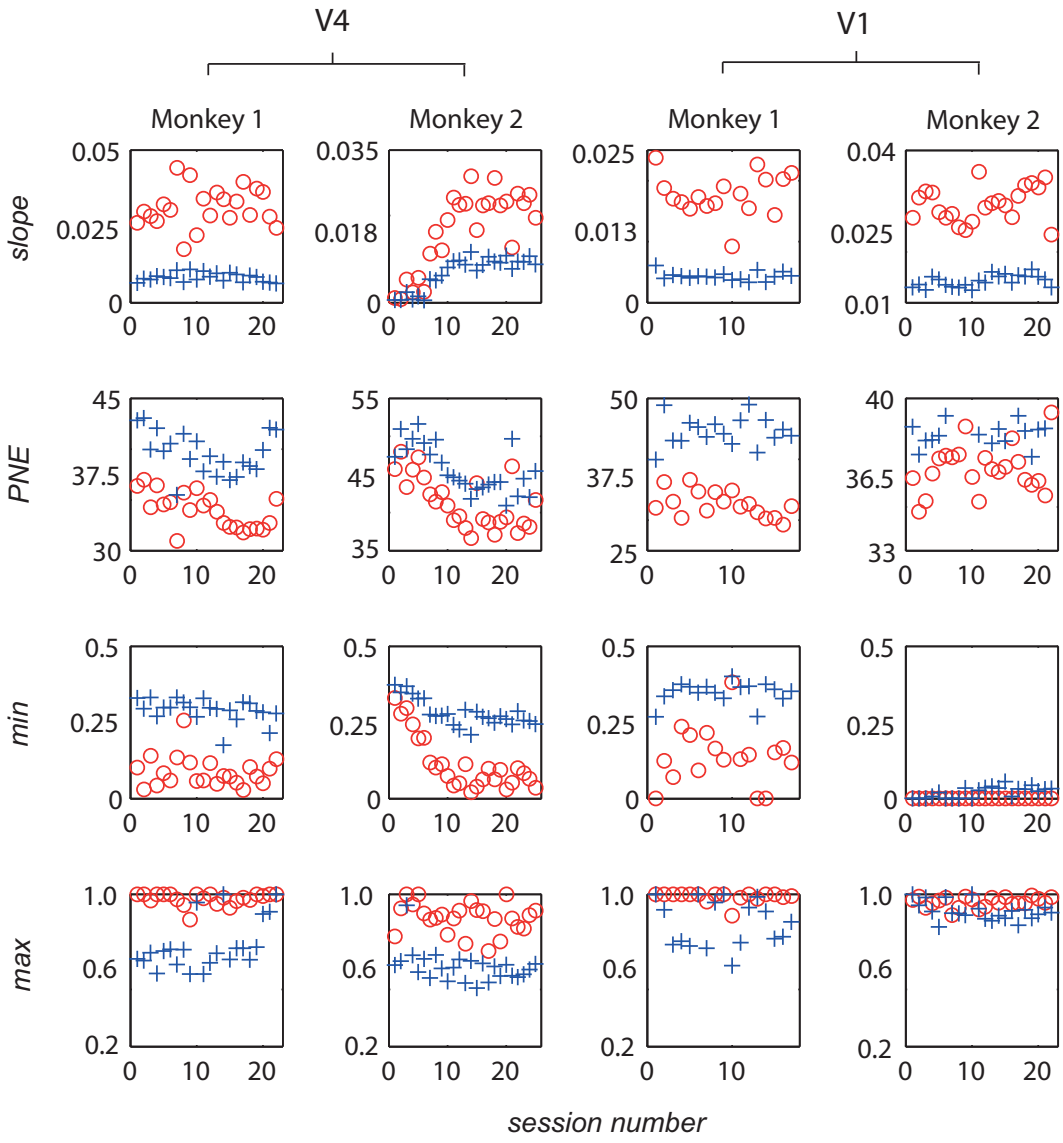


Figure 24. PROBMAT values were generated for population data using two distinct methods (blue crosses: *Pmean*; red circles: *Pcumulative*). The slope was consistently higher, and the PNE was consistently closer to the sample contrast, when PROBMAT was calculated based on a pooling of trial-wise activity across channels, than when it was generated separately for individual channels and then averaged together. Furthermore, the maxima tended to higher and the minima tended to be lower, with the *Pcumulative* method.

Values for these parameters were compared between the two methods using a *t*-test. As predicted, the cumulative method yielded better results across all four parameters, in both monkeys at the two recording sites (Table 8). The *Pcumulative* functions had steeper slopes, the PNEs were closer to the sample contrast, the minima were lower and the maxima were higher.

	t-test					
Statistic	*df*	*t*	*p*	*df*	*t*	*p*
	V4 location					
	Monkey 1			Monkey 2		
Slope	21	18.59	< .001*	24	8.69	< .001*
PNE	21	-19.59	< .001*	24	-12.27	< .001*
Min	21	-17.34	< .001*	24	-16.62	< .001*
Max	21	9.26	< .001*	24	16.86	< .001*
	V1 location					
	Monkey 1			Monkey 2		
Slope	16	18.13	< .001*	21	24.04	< .001*
PNE	16	-18.75	< .001*	21	-13.82	< .001*
Min	16	-11.36	< .001*	21	-5.46	< .001*
Max	16	5.23	< .001*	21	4.11	< .001*

* $q < \alpha$

Table 8. Results from a paired *t*-test which compared two different methods of calculating population PROBMAT values. In both monkeys and at both locations, *Pcumulative* values yielded better results than *Pmean* values, indicating that the pooling of activity across a population of neurons allowed higher-fidelity encoding of stimulus properties, than merely taking the mean of the individually fitted parameter values across single channels. An FDR correction was carried out for multiple comparisons (slope: $\alpha = .05/4\times4 = .05$; PNE: $\alpha = .05/4\times4 = .05$; minimum: $\alpha = .05/4\times4 = .05$; maximum: $\alpha = .05/4\times4 = .0125$).

Furthermore, an examination of changes in the slope and the minimum, at the V4 location in monkey 2, showed that for each of the parameters, values initially started at around the same levels during early sessions, but diverged between methods of PROBMAT calculation as training proceeded. This indicates that learning-induced alterations may have occurred in the pooling of responses across the population.

Based on these findings, PROBMAT values for all subsequent analyses of population data were thus calculated as *Pcumulative*.

1.6.4.2 Neurometric functions from population AUROC and PROBMAT values

Data were then pooled across all channels, in order to generate population AUROC and PROBMAT values. As stated previously, population AUROC values discard any information that may be present in trial-wise correlation of activity between sample and test responses, while this information is retained in PROBMAT data. Pooling methods were otherwise identical.

Population AUROC and PROBMAT values were plotted against test contrast, generating a pair of neurometric functions for each session (refer to Figure 25, Figure 26, Figure 27, and Figure 28 for population data). An examination of the data by eye indicated that the PROBMAT method resulted in consistently larger ranges of the neurometric function.

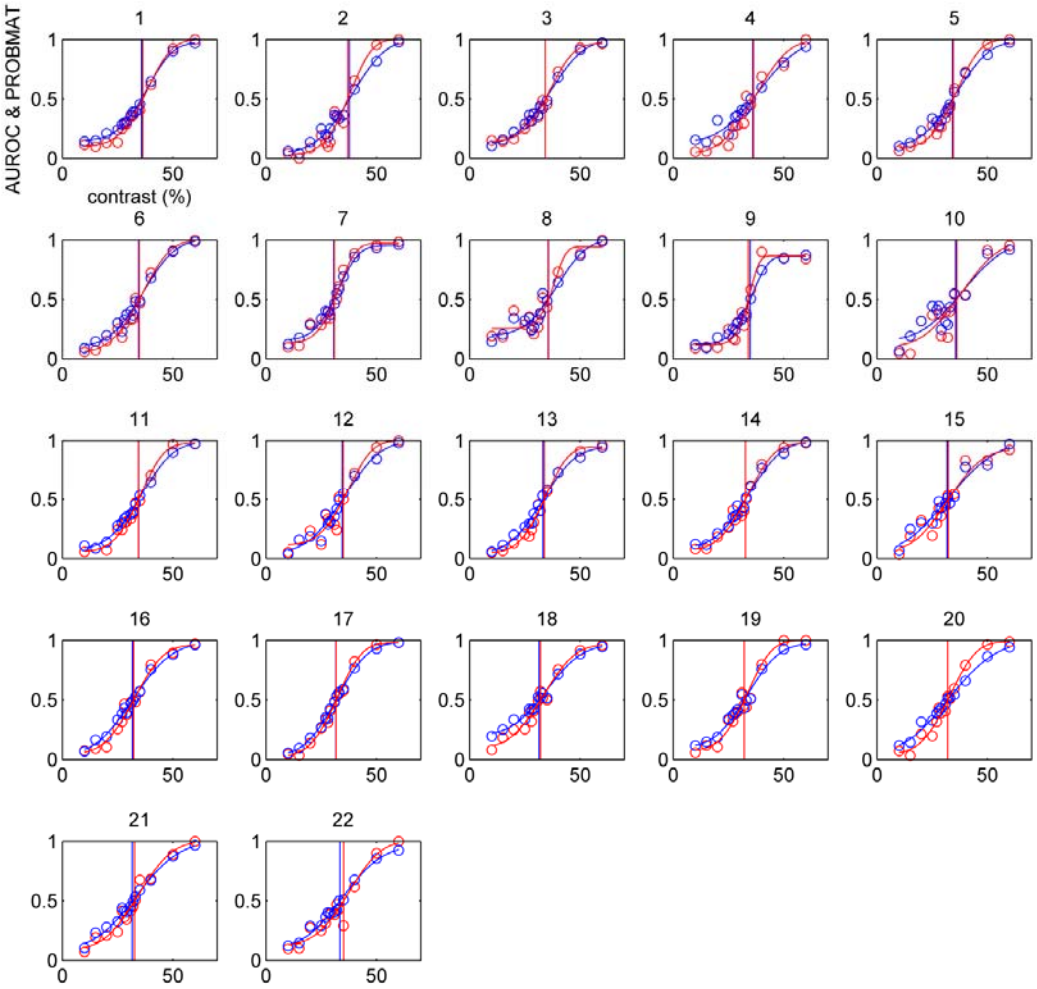


Figure 25. Neurometric functions of population AUROC (blue) and PROBMAT (red) values against test contrast, based on activity that was pooled across channels (monkey 1, V4 location). Each subplot presents data from one session. PROBMAT values tended to occupy a slightly wider range than AUROC values, indicating that trial-wise correlations do affect the decoding of neuronal activity. Thus, PROBMAT allowed a finer extraction of contrast-dependent information from spiking activity (Table 9). The x-axis corresponds to the test contrast, while the y-axis corresponds to AUROC and PROBMAT values.

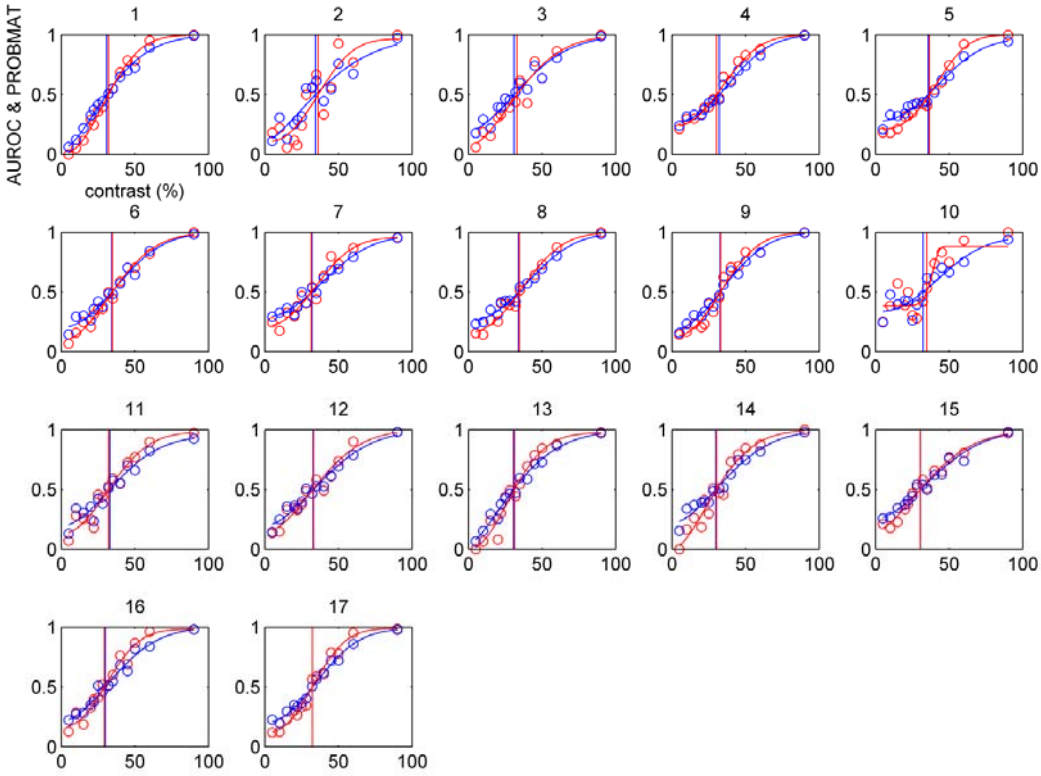


Figure 26. Neurometric functions of population AUROC (blue) and PROBMAT (red) values against test contrast (monkey 1, V1 location). The x-axis corresponds to the test contrast, while the y-axis corresponds to AUROC and PROBMAT values.

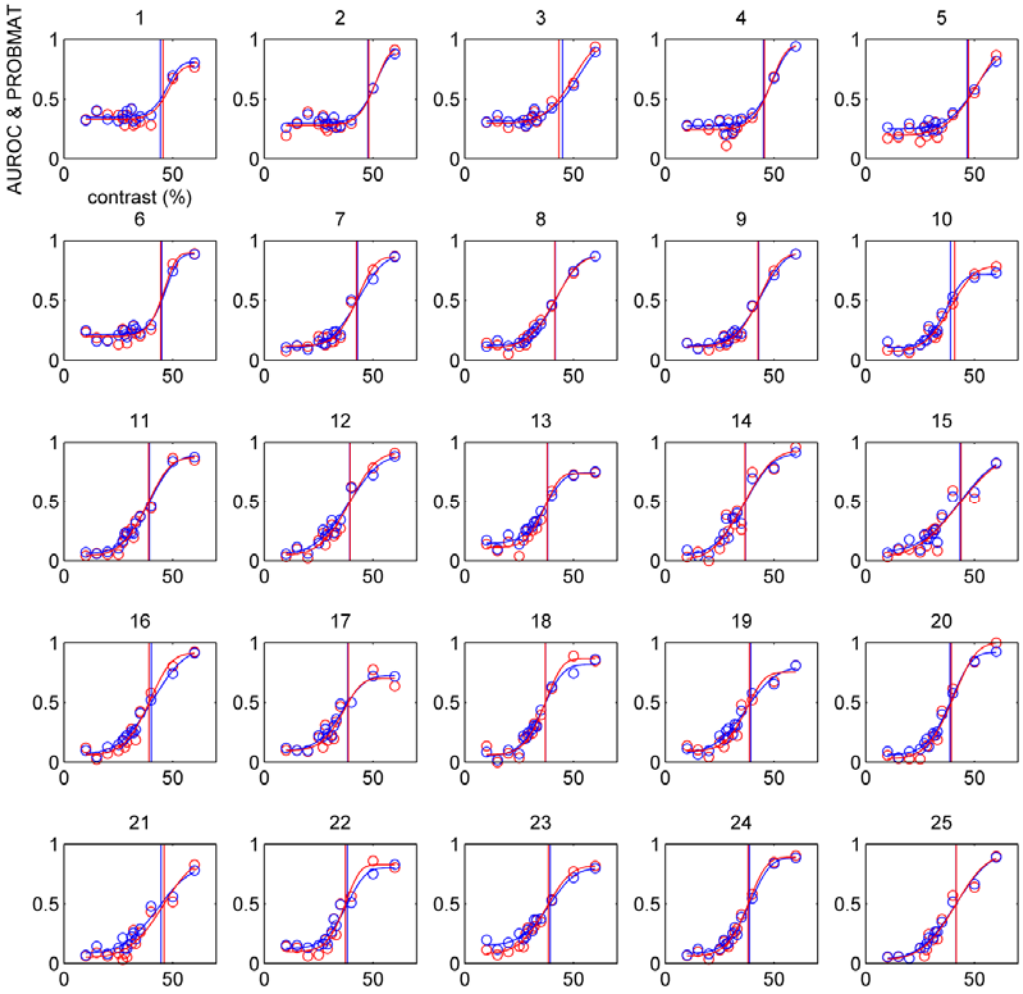


Figure 27. Neurometric functions of population AUROC (blue) and PROBMAT (red) values against test contrast (monkey 2, V4 location). The x-axis corresponds to the test contrast, while the y-axis corresponds to AUROC and PROBMAT values.

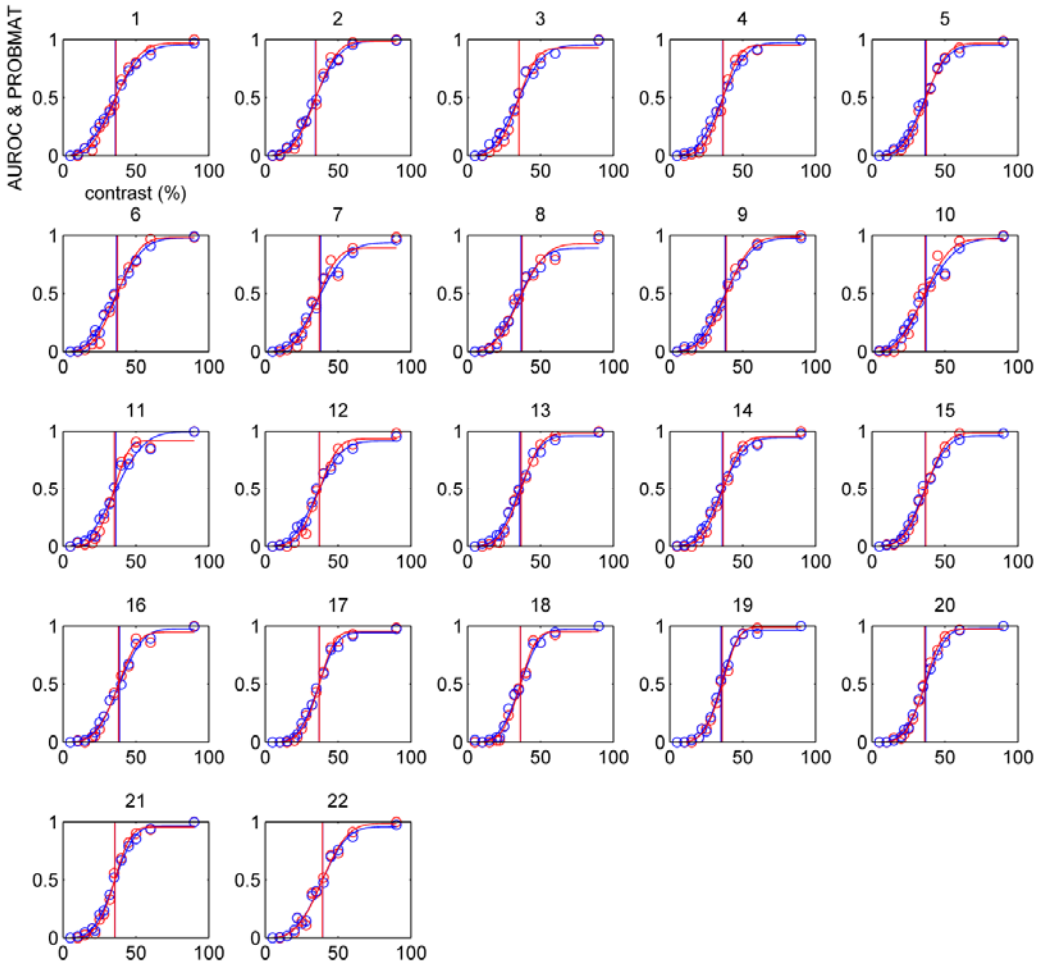


Figure 28. Neurometric functions of population AUROC (blue) and PROBMAT (red) values against test contrast (monkey 2, V1 location). The x-axis corresponds to the test contrast, while the y-axis corresponds to AUROC and PROBMAT values.

1.6.4.3 A comparison of population AUROC and PROBMAT values

As mentioned in a previous section on individual channel data, a comparison of the four parameters of the neurometric function was made between the two methods, to provide a quantitative measure of the degree of improvement offered by PROBMAT (Figure 29). A paired *t*-test was carried out to compare values between the two methods, for each parameter, and an FDR correction for multiple comparisons was applied to alpha levels.

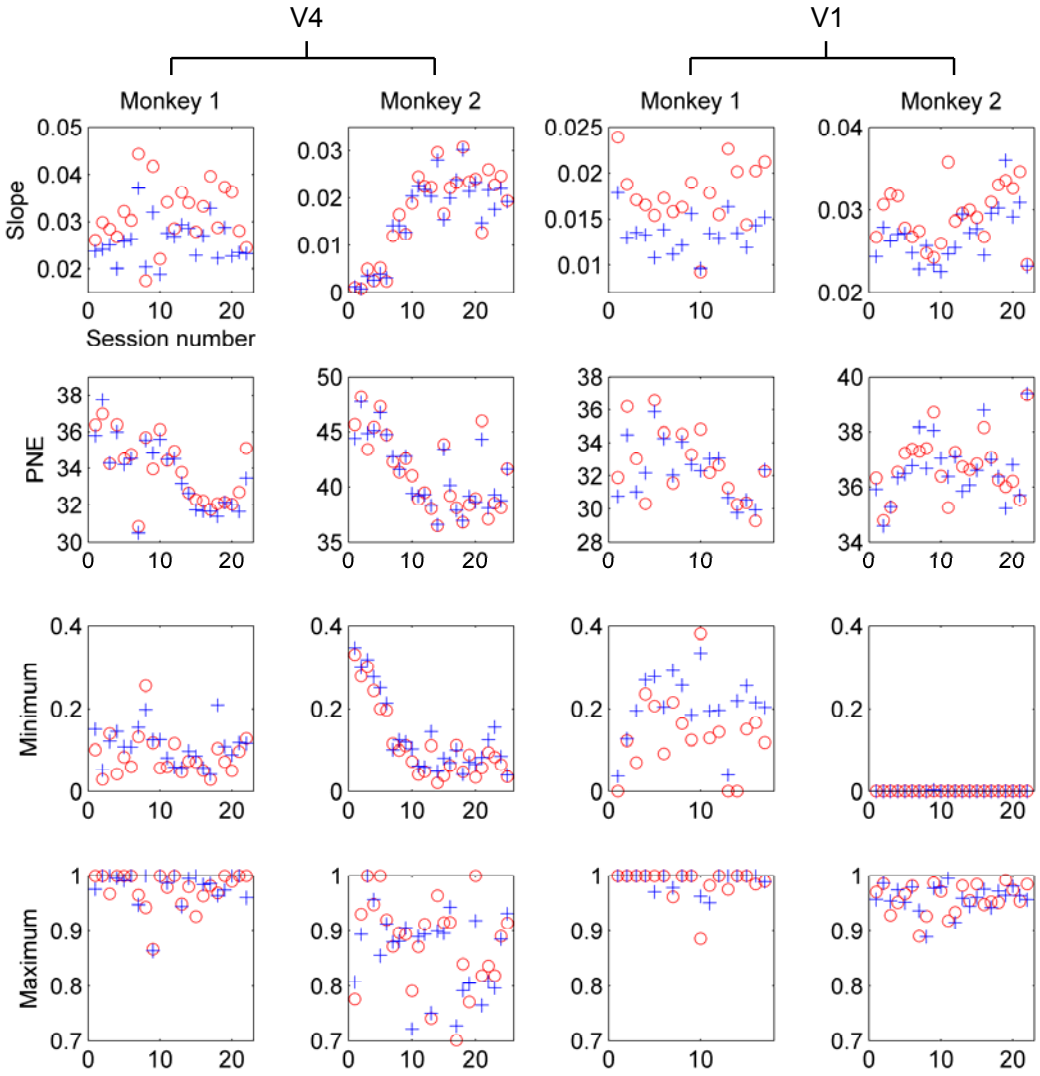


Figure 29. Parameter values of the psychometric function against training session. First and second columns: V4; third and fourth columns: V1. First and third columns: monkey 1; second and fourth columns: monkey 2. First row: slope; second row: PNE; third row: minimum value; fourth row: maximum value. Blue 'plus' symbols: AUROC values; red circles: PROBMAT values.

For both of the subjects and at both recording sites, the slope of the psychometric function obtained using the PROBMAT approach was steeper than that obtained using the AUROC method (paired *t*-test, Table 9). Furthermore, the minima were reduced for V4 recording sites in both monkeys and for the V1 recording site in monkey 1. Thus, the use of within-trial comparisons of activity allowed an ideal observer to be more accurate when discriminating between sample and test stimuli.

	t-test					
	df	*t*	*q*	*df*	*t*	*q*
	V4 location					
	Monkey 1			Monkey 2		
Slope	21	7.7	< .001*	24	2.54	.0178*
PNE	21	2.61	.0162	24	-0.05	.959
Min	21	-2.6	.0166*	24	-4.83	< .001*
Max	21	-0.73	.472	24	2.0	.0564
	V1 location					
	Monkey 1			Monkey 2		
Slope	16	9.8	< .001*	21	4.45	< .001*
PNE	16	1.35	.196	21	0.77	.448
Min	16	-5.06	< .001*	21	-1.0	.329
Max	16	-0.75	.467	21	-0.14	.893

* $q < \alpha$

Table 9. Results from a paired *t*-test, comparing values of each of the parameters derived from AUROC and PROBMAT methods. The PROBMAT approach yielded higher values for the slope of the curve at 30% contrast, for both monkeys and both recording locations (slope: $\alpha = .05/4 \times 4 = .05$; PNE: $\alpha = .05/4 = .0125$; minimum: $\alpha = .05/4 \times 3 = .0375$; maximum: $\alpha = .05/4 = .0125$; an FDR correction was carried out as described in the section, 'Corrections for multiple comparisons,' on page 27). The minimum values produced by the trial-wise method were also significantly lower for both subjects at the V4 location, and for monkey 1 at the V1 location.

Due to the advantage conferred by PROBMAT (under the conditions of the current study), changes in the neurometric function with training were evaluated using data derived through this method.

1.6.4.4 Changes of the population neurometric function with training

As previously described, sets of population PROBMAT values were calculated for each session by combining data across all channels, for each trial and each test contrast condition (Figure 25, Figure 26, Figure 27, and Figure 28). To identify changes in the neurometric function with time, a Spearman's rank correlation analysis was calculated between the parameters of interest and session number (Table 10).

Over the course of training, the population PNE in V4 decreased significantly from around 36% to 33% for monkey 1, and from around 45% to 38% for monkey 2. A significant increase in slope was also observed for monkey 2 at the V4 location. This observation mirrored the changes in the slope of the CRF, reported in the section,

'*Contrast response functions*,' page 57, in which the population CRF became significantly steeper around the contrast of 30% in monkey 2, but not in monkey 1.

At the V1 location, changes were less evident than at the V4 location. The PNE appeared to shift towards 30% for monkey 1, but this trend was not significant. No changes in the other parameters were observed.

	Spearman's rank correlation					
	df	*r*	*q*	*df*	*r*	*q*
	V4 location					
	Monkey 1			Monkey 2		
Slope	20	.171	.445	23	.752	< .001*
PNE	20	-.591	.00444*	23	-.612	.00115*
Min	20	-.073	.747	23	-.708	< .001*
Max	20	-.019	.936	23	-.297	.149
	V1 location					
	Monkey 1			Monkey 2		
Slope	15	.1	.701	20	.265	.233
PNE	15	-.48	.053	20	.077	.732
Min	15	.047	.861	20	.442	.0393
Max	15	-.485	.0503	20	.171	.445

* $q < \alpha$

Table 10. Changes in population neurometric functions with training. The PNE for each population of V4 neurons shifted significantly towards the left in both subjects, towards the value of 30%. A significant increase in slope, as well as a decrease in the minimum value, was also observed for recordings at the V4 location in monkey 2 (Spearman's rank correlation, FDR correction, $\alpha = .05/16 \times 4 = .0125$).

1.6.4.5 Contrast-dependent PROBMAT as a function of time

To visualise changes in PROBMAT values for each condition as a function of time, population PROBMAT values were plotted against session number, for each test contrast condition, and an exponential function was fit to the data (Figure 30). Improvements over the course of training were particularly pronounced for monkey 2, at the V4 location, and the pattern of change was similar to that seen at the behavioural level, in the psychophysical data (presented in the section, '*Perceptual learning for individual test contrast conditions*,' Figure 30, page 81).

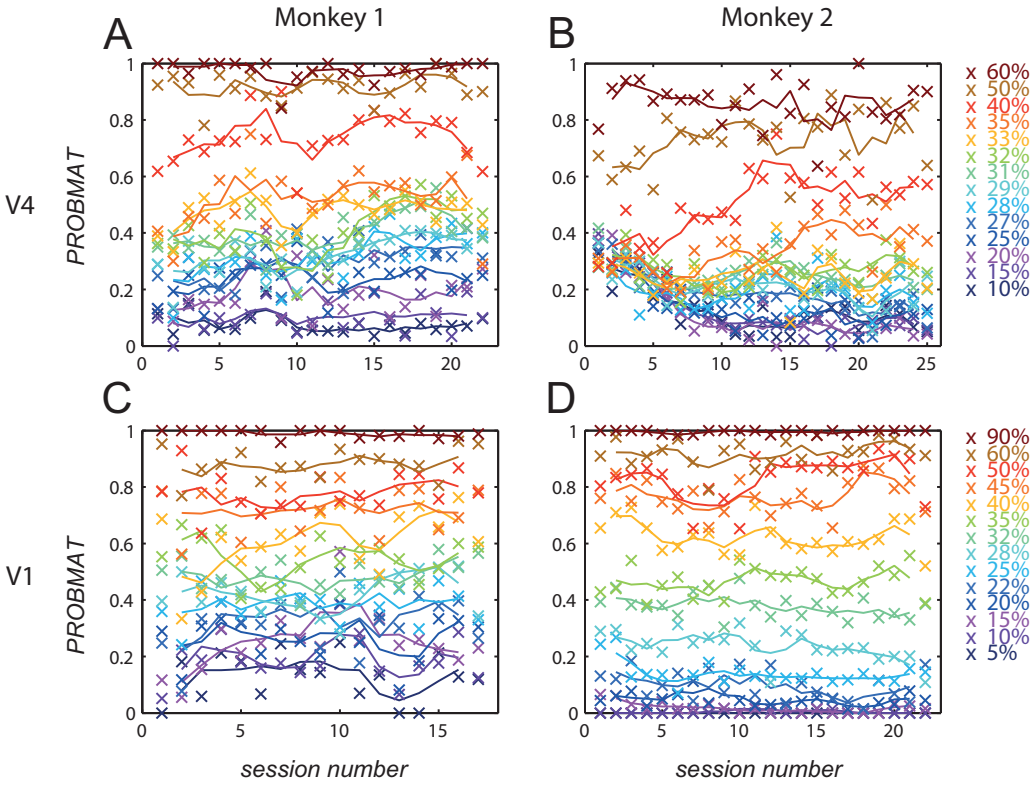


Figure 30. Population PROBMAT values were plotted against session number, for each test contrast condition. A & B: V4; C & D: V1. A & C: monkey 1; B & D: monkey 2. Changes were particularly pronounced when monkey 2 was trained with stimuli at the V4 location, during conditions with low test contrasts; this pattern mirrored that seen in the behavioural data. Lines represent the running average, calculated across three consecutive sessions at a time.

1.6.5 Exclusion of channels with stimulus-evoked suppression

In four channels, neurons showed stimulus-evoked inhibition, rather than excitation (as reported in the section, '*Exclusion of channels with stimulus-evoked suppression*,' page 60). For these channels, PROBMAT values were fit with a Weibull function that had a negative slope (i.e. the plots were flipped about the axis $y = 0.5$). The analysis was then repeated with the exclusion of these four channels (note that as none of the V1 channels exhibited stimulus-evoked suppression, this analysis was only carried out on V4 data).

Their exclusion did not affect the results obtained for changes in the population neurometric function, as reported in the previous section. The four stimulus-suppressed channels were thus included in all subsequent analyses.

1.6.6 Effects of data normalisation

Thus far, analyses of population PROBMAT values were carried out on data that was combined across channels through a simple summation of the firing rate across channels. This meant that channels with higher absolute firing rates made a proportionately larger contribution to the population data than did channels with weaker responses. One might hypothesise that the differences in maximum firing rates between different channel recordings reflected meaningful differences in the contributions of individual neurons to the interpretation of stimuli.

On the other hand, it was possible that some form of normalisation occurred at a later stage of decoding in the visual processing hierarchy, effectively leading to a reassignment of weights. The contributions of neurons with high firing rates may have been reduced, while those of neurons with less activity may have been boosted. To test this theory, the activity of each channel was normalised to its maximum levels prior to pooling; each channel was thus assigned an equal weight. A paired *t*-test was used to compare parameter values of the PROBMAT functions that were generated in the presence or absence of normalisation.

The comparison revealed that the normalisation of single-channel activity yielded poorer results. After normalisation, the slope of the neurometric function was shallower and the minimum value was higher, for V4 recordings in monkey 2, and for V1 recordings in monkey 1 (Table 11). In other words, the ranges of PROBMAT values were significantly reduced when the contributions of individual channels were equalised. A significant difference was also seen in the PNE- normalisation lowered the PNE towards 30%, for monkey 1 at the V4 location and monkey 2 at the V1 location, while it raised the PNE away from 30%, for monkey 2 at the V4 location.

	descriptive statistics					*t*-test	
	M1	*SD1*	*M2*	*SD2*	*df*	*t*	*q*
				V4 location			
				Monkey 1			
slope	0.0370	0.0071	0.0374	0.0069	21	0.54	.593
PNE	33.0	1.6	32.2	1.4	21	-8.12	< .001*
min	0.05	0.04	0.05	0.03	21	0.30	.769
max	0.98	0.03	0.98	0.03	21	0.07	.941
				Monkey 2			
slope	0.0181	0.0099	0.0165	0.0091	24	-4.86	< .001*
PNE	40.5	3.4	41.0	3.3	24	2.85	.00880*
min	0.12	0.10	0.14	0.10	24	5.63	< .001*
max	0.88	0.08	0.87	0.08	24	-0.86	.401
				V1 location			
				Monkey 1			
slope	0.0177	0.0035	0.0158	0.0026	16	-3.85	.00141*
PNE	32.7	2.1	32.5	2.1	16	-0.70	.491
min	0.14	0.10	0.19	0.10	16	3.63	.00223*
max	0.99	0.03	0.99	0.03	16	1.40	.181
				Monkey 2			
slope	0.0294	0.0034	0.0297	0.0034	21	0.87	.392
PNE	36.8	1.1	36.3	1.2	21	-4.52	< .001*
min	0.00	0.00	0.00	0.00	21	1.00	.329
max	0.96	0.03	0.96	0.03	21	-0.34	.740

* $q < \alpha$

Table 11. A comparison of population results, before (*M1*) and after (*M2*) normalisation of data to the maximum responses of individual channels. The slope of the neurometric function decreased, and the minimum value increased after normalisation, for V4 responses in monkey 2 and for V1 responses in monkey 1, indicating that normalisation made the 'readout' of population data slightly worse. Effects of normalisation on the PNE were not consistent across different recording sites.

1.6.7 Within-trial single-channel correlations in spiking activity

As stated above, the superiority of the PROBMAT method over the traditional ROC method likely stemmed from the fact it took within-trial activity correlations into account, while the benefit of pooling across neurons might be limited by the potential presence of noise correlations between channels. While both factors rely on the existence of correlations in firing rate from trial to trial, they nevertheless make distinct contributions to the workings of a hypothetical decoder. If within-trial correlations in sample-evoked and test-evoked activity were perfect (i.e. yielding a correlation

coefficient of one), then even if noise correlations were present between channels, they would not affect levels of information carried by the signal. Similarly, if within-trial correlations were completely absent, then noise correlations would not affect the amount of information content. However, if within-trial correlations were smaller than one but larger than zero, then the presence of positive noise correlations between channels would depend on the level of signal correlation that was present between channels (Averbeck, Latham, & Pouget, 2006). Equivalent reasoning would apply to negative correlations in within-trial single-channel activity and their respective noise correlations.

We first investigated the degree of within-trial correlation by calculating the activity that was elicited by the sample stimulus and the activity that was elicited by the test stimulus. To enable comparisons between different test contrasts, the activity related to the sample was converted to a *z*-score (combined across all responses within a session), according to the formula:

$$Z = \frac{X - M}{SD} \qquad \text{… (Equation 11),}$$

where X represents the measured single-trial activity in response to the sample; M corresponds to the mean of single-trial activities for a given sample contrast; and SD is the standard deviation of single-trial activities for a given sample contrast. The levels of activity elicited by a given test contrast were converted to a *z*-score in a similar manner, thereby removing contrast-dependent signal correlations from the analysis. To provide an initial overview of the data, all the *z*-scored values were then pooled across sessions for individual channels, and an 'across-session-within-trial' correlation coefficient, R_w, was calculated for each channel (data from two example channels are shown in Figure 31). Correlation coefficients were significant and positive for all channels ($p < .001$, Pearson's correlation). The distribution of correlation coefficients for each of the two monkeys and areas is shown in Figure 32.

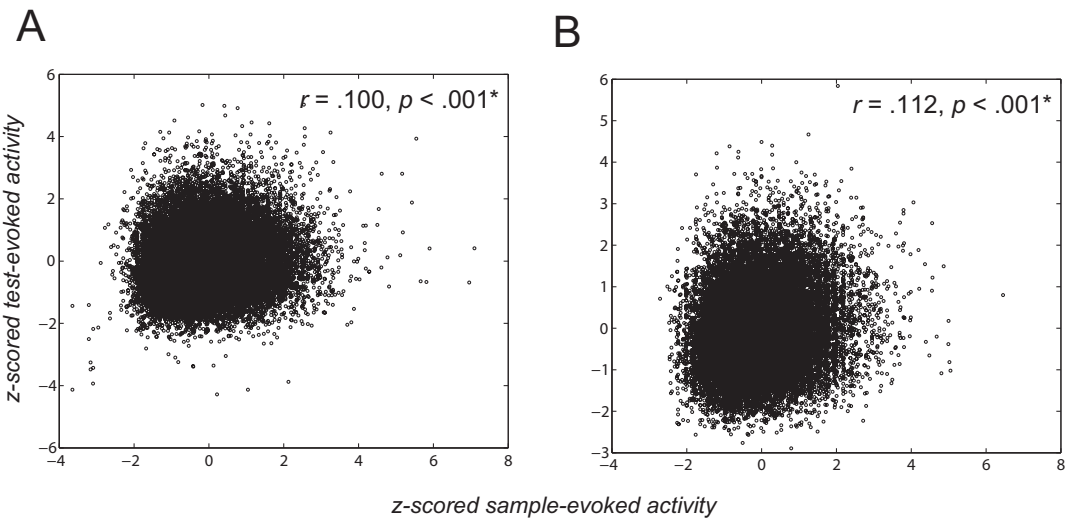


Figure 31. Correlations between sample- and test-evoked activity, across all training sessions, for two example channels (A: channel 18, monkey 2, V1 location; B: channel 20, monkey 1, V4 location). Activity levels were converted to *z*-scores for each stimulus contrast and day, prior to the calculation of correlations.

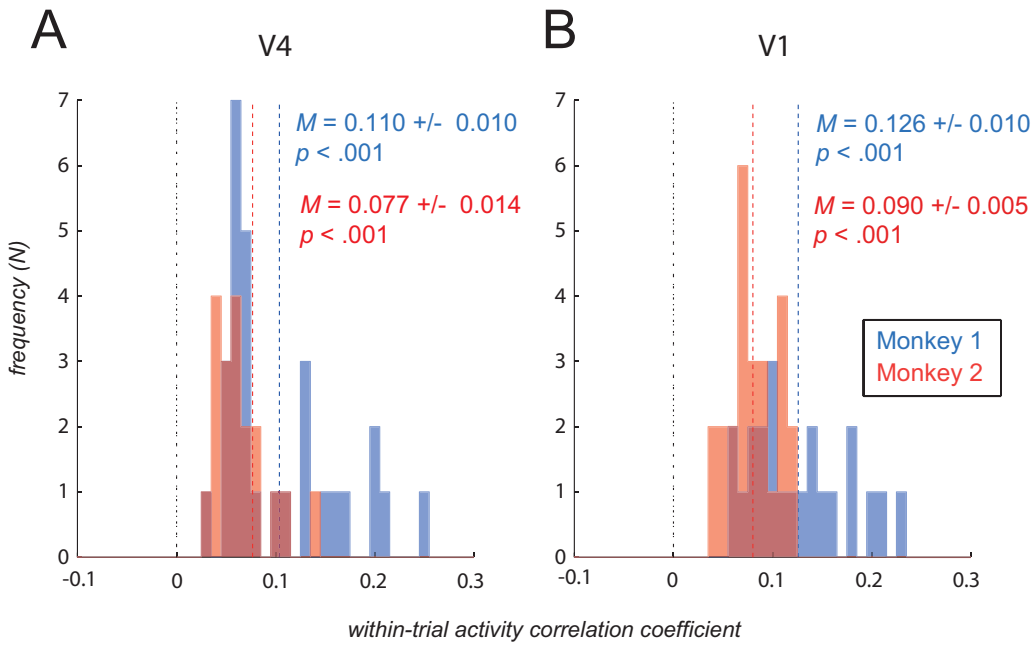


Figure 32. Distributions of correlation coefficients for sample-versus-test within-trial activity for the two monkeys (blue: monkey 1; red: monkey 2) and recording areas. A: V4 location; B: V1 location. A *t*-test indicated that distributions were significantly different from zero. Error indicates 1 SEM. The vertical black dotted line demarcates within-trial activity R-values of 0; the blue and red vertical dotted lines indicate the means of the distributions for monkeys 1 and 2 respectively.

Next, to examine whether within-trial correlations changed with training, a correlation was calculated between R_w and session number. In V4, 9 channels showed significant changes with training. In each of these cases, the relationship between the coefficients and the session number was positive, suggesting that within-trial correlations might have increased with training. Five of these channels were recorded in monkey 1, while 4 were recorded in monkey 2. In V1, significant changes in within-trial correlations over the course of training were seen in 6 channels. Five of these channels (all of which were from monkey 2) showed negative correlations, while one channel (from monkey 1) showed a positive correlation. These changes are depicted for the different areas and for the different monkeys in Figure 33A.

We also investigated whether changes in within-trial correlations in activity were present in each area when all the channels were included (rather than when only significant ones were included). When all the channels were taken into account, significant negative changes were only present for area V1 in monkey 2 (see Figure 33B), while no changes were found in the other areas.

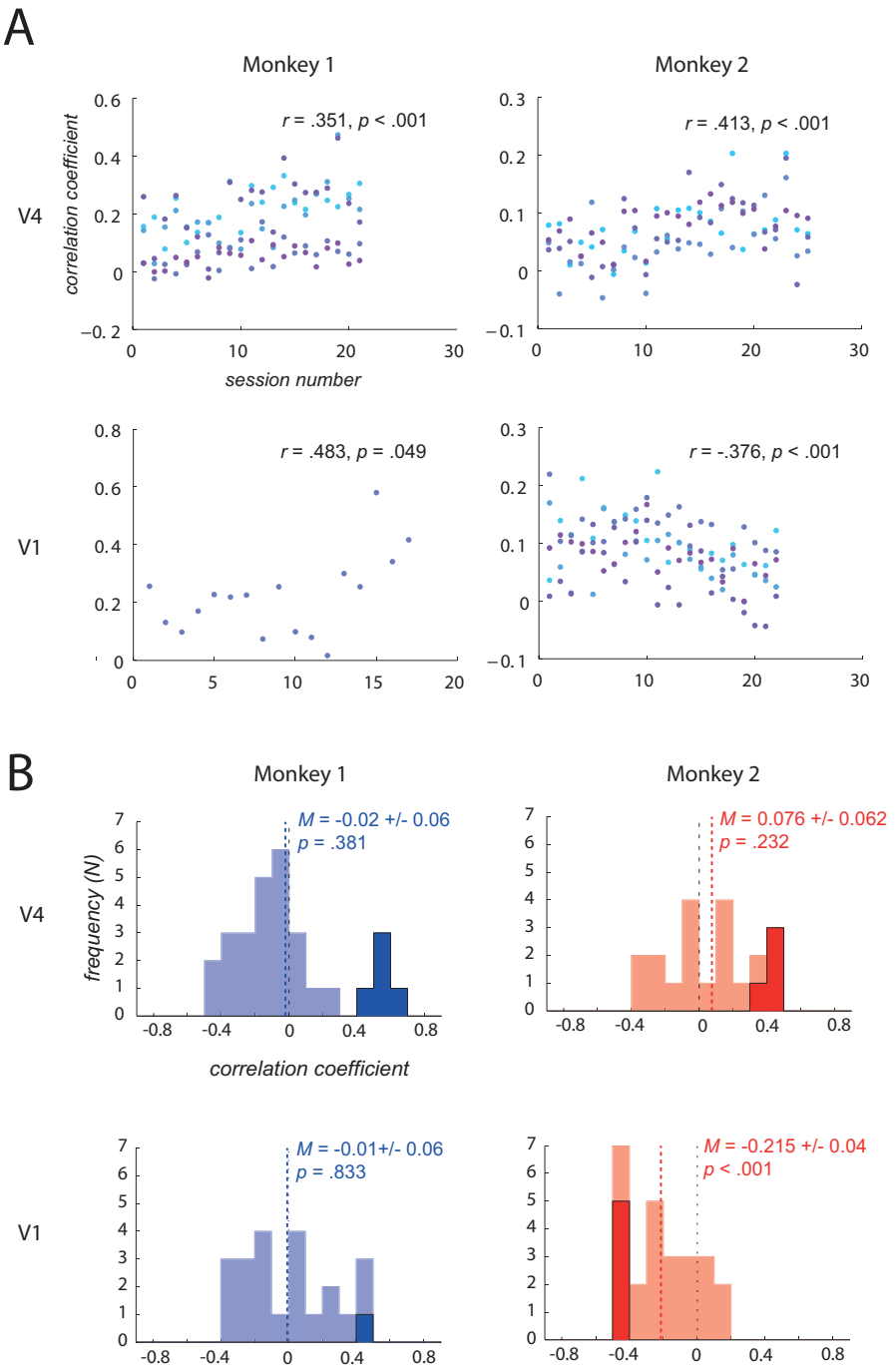


Figure 33. Changes in within-trial correlations of activity with training. A) Correlation coefficients of within-trial activity, R_w, between sample and test responses, as a function of time, for only those recording channels where significant changes occurred with training. Data from individual channels are coded by colour, for each recording site. Values of *r* and *p* indicate correlations across significant channels, for each of the recording sites. B) Distributions of correlation coefficients across all channels (regardless of whether significant changes occurred with training), from each recording site. Dark shaded histograms indicate channels for which significant changes were seen;

light shaded histograms indicate the distribution of correlation coefficients for channels that did not show significant changes. Error values correspond to 1 SEM; *p*-values indicate whether the means of the distributions differed significantly from zero. Dashed vertical lines indicate the location of the means.

1.6.8 PROBMAT and noise correlations

As previously stated, the PROBMAT technique was likely to aid decoding of single trial activity, due to the existence of within-trial correlations in activity between responses to the sample and test stimuli; however, a substantial contribution (or impediment) to the decoding of such within-trial population activity levels may also have arisen due to the existence of noise correlations between neurons, i.e. co-fluctuations of activity levels between channels. If noise correlations tended to be positive, then that could limit the level of within-trial activity correlation seen at the population level, and hence limit the decoding of population activity (but as stated above, this would depend on the level of signal correlations that were present (Averbeck et al., 2006)).

Thus, an analysis was carried out to explore the degree of noise correlation that was present in our data and to determine whether this changed as learning progressed. For each day, trial, and channel, the sample-evoked activity was calculated, and these data were converted into *z*-scores to remove signal-dependent correlations. Noise correlations (represented by Pearson's correlation coefficients) were then calculated between all possible channel combinations, for each recording day. Distributions of correlation coefficients were compared between early sessions and late sessions, using various criteria to allocate the sessions into early and late groups. Figure 34 displays the histogram distributions of correlation coefficients for a comparison involving the first and the last five training days.

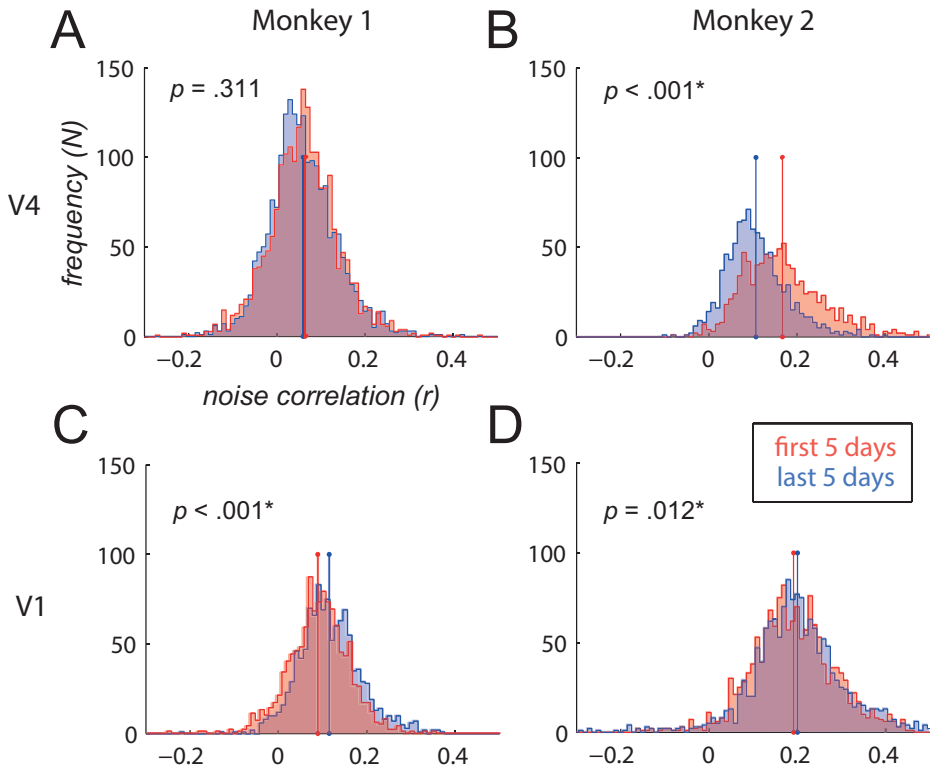


Figure 34. Distributions of noise correlation coefficients for the first (red) and the last (blue) five days of training in monkey 1 (A & C) and monkey 2 (B & D). A & B: V4 location; C & D: V1 location. Solid vertical lines show the means of the distributions. *P*-values indicated whether the means of the distributions differed significantly between early and late sessions (Student's *t*-test).

On average, noise correlations were larger in V1 than in V4 (see Figure 34). Noise correlations increased significantly in both monkeys for V1 channels ($p < .05$, Student's *t*-test), while they decreased in both monkeys for V4 channels (though this was significant only for monkey 2, $p < .001$, *t*-test). This pattern of results was present regardless of the number of sessions included within the early and late groups (i.e. a variety of groupings were used and tested, each of which yielded a significant effect in both monkeys and both areas).

1.6.9 Neurometric versus psychometric thresholds

Thresholds at 82% and 18% neurometric performance were taken from population PROBMAT-versus-contrast functions and monitored over time for training-induced changes. (Single-channel PROBMAT data often did not yield fitted functions

that reached these threshold levels, thus this analysis was carried out solely on population PROBMAT data.) Two thresholds were obtained for the following reasons:

1. The threshold value, T_L, was calculated for conditions where the contrast of the test stimulus was lower than that of the sample, and provided a measure of the discriminability of spiking activity to low-contrast test stimuli. T_L corresponded to the contrast at which PROBMAT was equal to 18%.
2. The threshold value, T_H, was calculated for conditions where the contrast of the test stimulus was higher than that of the sample, and provided a measure of the discriminability of spiking activity to high contrast test stimuli. T_H corresponded to the contrast at which PROBMAT was equal to 82%.

As with the psychometric data, thresholds could not be obtained during some sessions. For these sessions, the threshold was assigned the highest possible value (T_L = 30% for C_L conditions; $T_H = 100 - 30 = 70\%$ for C_H conditions), and data points for these sessions were indicated by an unfilled circle (see Figure 35).

A Spearman's rank correlation was performed to identify changes in thresholds with time. Improvements in lower neurometric thresholds occurred in monkey 2, at both locations (refer to Table 12 for results of the correlation analysis), matching the decreases in psychometric threshold seen in the behavioural data.

Statistic	*df*	*r*	*q*	*df*	*r*	*q*
	Monkey 1			Monkey 2		
			V4			
C_L	20	.487	.0227	23	-.549	.00444*
C_H	20	-.178	.427	23	-.177	.397
			V1			
C_L	15	.111	.672	20	-.688	< .001*
C_H	15	-.210	.419	20	-.311	.159

* $q < \alpha$

Table 12. Spearman's rank correlation coefficients and *q*-values, from an examination of changes in neurometric and psychometric thresholds over the course of training with non-roving stimuli. FDR correction, $\alpha = .05/8 \times 2 = .0125$.

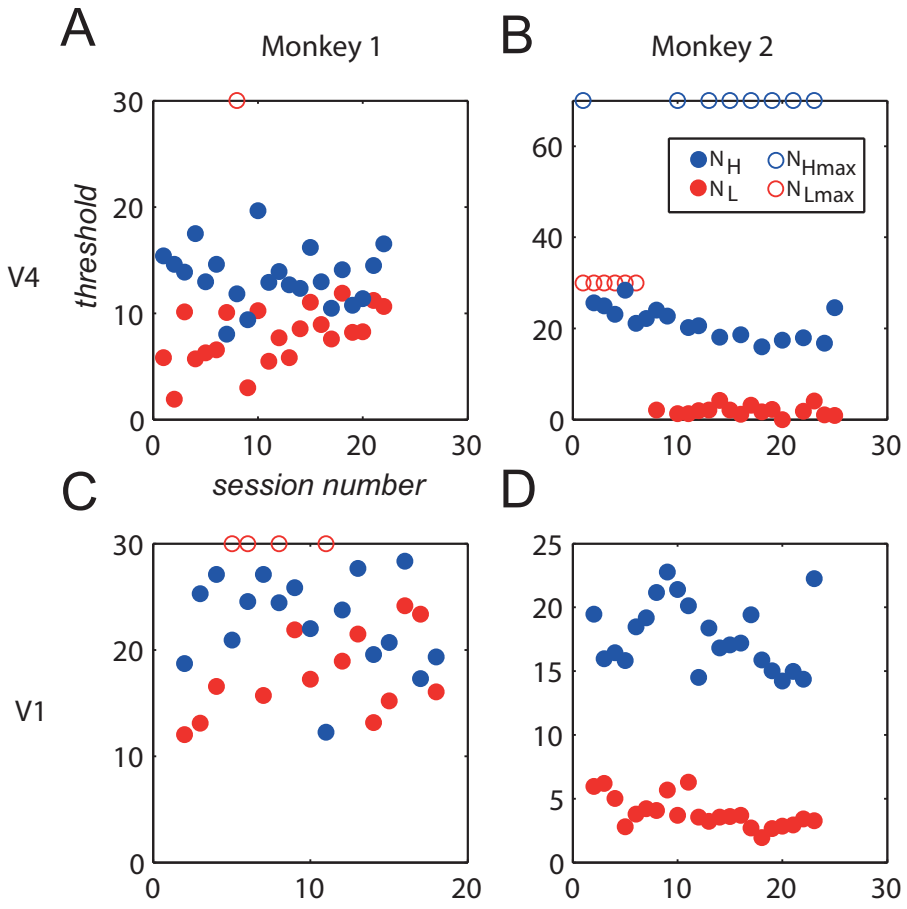


Figure 35. Neurometric thresholds as a function of training session. A & B: V4; C & D: V1. A & C: monkey 1; B & D: monkey 2. Filled markers: actual neurometric threshold values; unfilled markers: threshold values assigned as maximum levels. Red markers: N_L conditions (the test contrast was lower than that of the sample); blue markers: N_H conditions (the test contrast was higher than that of the sample).

1.6.10 Effects of adaptation on stimulus-evoked activity

1.6.10.1 Effects of adaptation on responses to the test

Upon examination of the neuronal raster plots and PSTHs, it was noticed that for some channels, during conditions where the test contrast was slightly higher than the sample contrast, the test-induced response was nevertheless smaller than the sample-induced response (Figure 36). This reduction of firing activity to a higher-intensity stimulus was likely to have been due to adaptation.

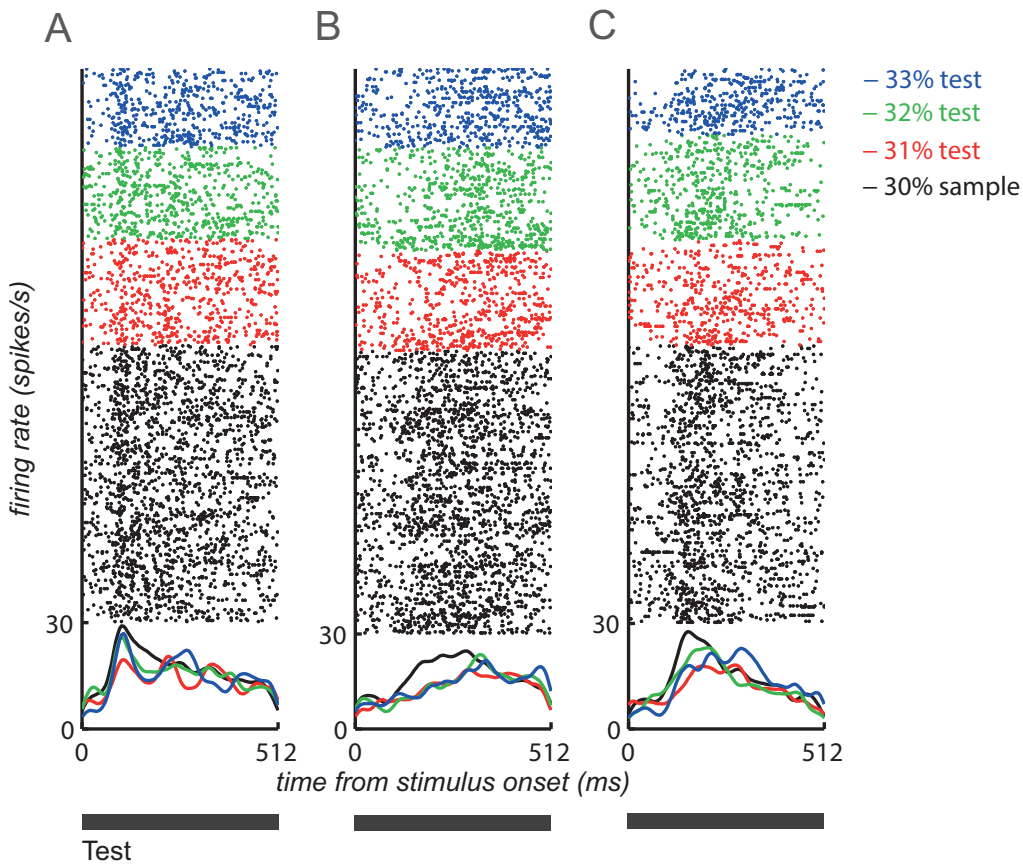


Figure 36. PSTHs of stimulus-evoked spiking activity from three example channels (monkey 2, V4 location, channels 10, 52 and 53 from sessions 77, 75 and 46, respectively). Peak activity levels elicited by the test stimuli (red: 31% contrast; green: 32%; blue: 33%) were lower than those evoked by the sample (black: 30%), even though the test contrast was higher than the sample contrast during each of these three conditions.

This phenomenon typically occurs when two stimuli are presented in close succession- after presentation of the first stimulus, the response elicited by the second stimulus is lower compared to what it would have been, had the first stimulus been absent. To examine the degree of contrast adaptation in our paradigm, firing rates were compared between sample and test stimulus presentations for each channel, for conditions where the test contrast was just above 30%.

Significant contrast adaptation was found for a subset of channels and conditions (paired *t*-test, FDR correction for multiple test contrasts, $q < \alpha$). For monkey 1, when stimuli were presented at the V4 location, significant differences were seen on the majority of channels (15/29 channels showed significantly higher responses to the

sample for at least one of the three test contrast conditions, see Table 13 for the breakdown by condition). For monkey 2, when stimuli were presented at the V4 location, significant differences were seen on all of the channels, and without exception, responses were lower to the test than to the sample (20/20 channels), despite the fact that the test stimulus was of higher contrast than the sample.

When stimuli were presented at the V1 location, results differed even more between the two subjects. For monkey 1, significant differences were seen in a few channels (2/23 channels had higher activity for the sample, while 1/23 had higher activity for the test). For monkey 2, significant differences were seen on a majority of channels (22/25 channels had higher activity for the sample, while 1/25 had higher activity for the test).

Area	Condition	Monkey 1		Monkey 2	
		Sample > Test	Test > Sample	Sample > Test	Test > Sample
V4	31%	14	9	20	0
	32%	13	9	19	0
	33%	8	11	19	0
V1	32%	2	1	21	1

Table 13. Number of channels where significant differences between test- and sample-induced activity occurred, when test and sample contrasts differed only slightly. For monkey 1, response adaptation was seen in around half of the V4 channels ($N = 29$) and in hardly any of the V1 channels ($N = 23$), whereas for monkey 2, adaptation occurred in the vast majority of V4 ($N = 20$) and V1 ($N = 25$) channels.

When data were combined across channels, the results obtained at the population level matched those seen for individual channels (Figure 37). Responses to the test stimulus were significantly lower for monkey 1, when stimuli were presented at the V4 location (31% test contrast: $t(58347) = -13.5$, $p < .001$; 32% test contrast: $t(51068) = -9.6$, $p < .001$; 33% test contrast: $t(45616) = -2.8$, $p = .005$), and a less pronounced but still significant effect of adaptation was seen for stimuli at the V1 location (32% test contrast: $t(30865) = -2.0$, $p = .0453$). For monkey 2, responses were lower during test stimulus presentations, at both locations (V4: 31% test contrast: $t(37919) = -48.7$, $p < .001$; 32% test contrast: $t(34019) = -39.3$, $p < .001$; 33% test contrast: $t(33319) = -41.9$, $p < .001$; V1: 32% test contrast: $t(55474) = -33.8$, $p < .001$, FDR correction, $\alpha = .05/8 \times 8 = .05$).

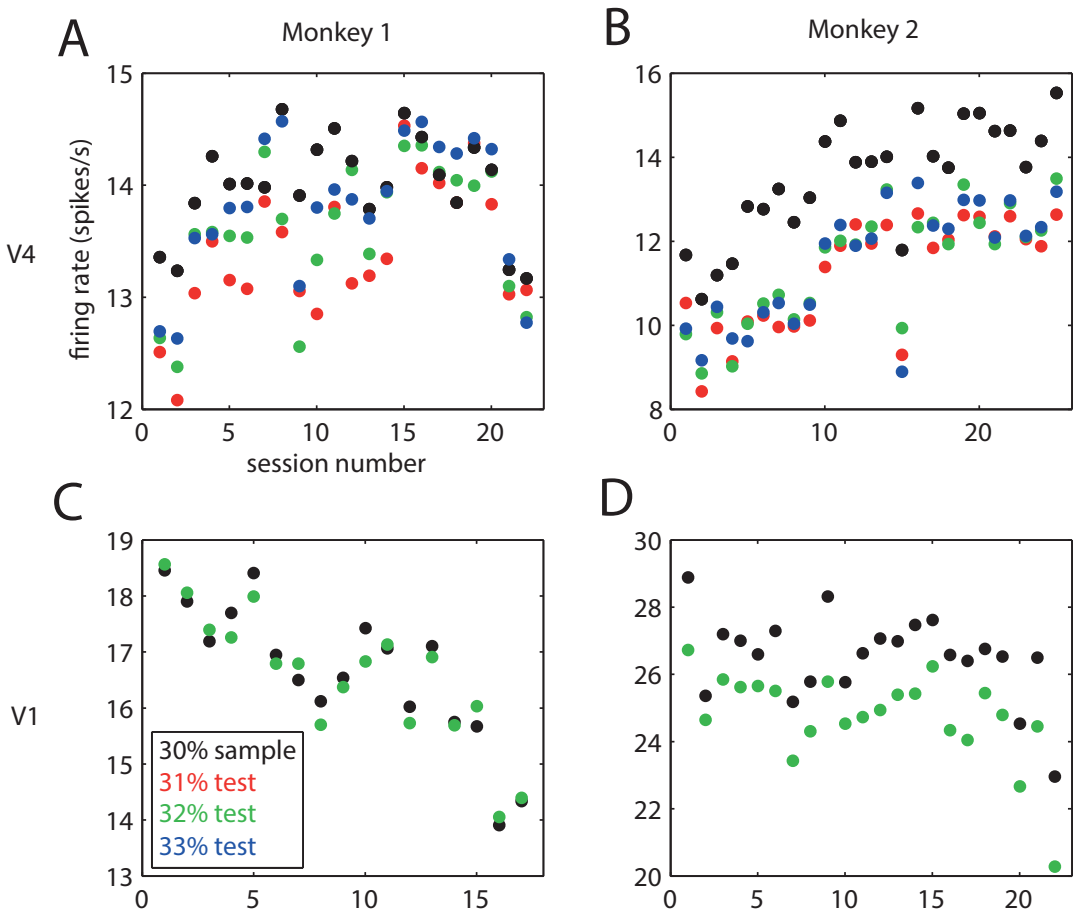


Figure 37. Plots of mean firing rates across channels against session number, to identify adaptation-related differences in stimulus-evoked activity during conditions where the test contrast was just above 30% (red: 31%; green: 32%; blue: 33%). A & B: V4; C & D: V1. A & C: monkey 1; B & D: monkey 2. Adaptation was visible in many cases (indicated by black markers that are located above coloured ones).

1.6.10.2 Changes in adaptation with training

To investigate whether the effects of adaptation changed over the course of training, an adaptation index, *AI*, was calculated, and values of *AI* were plotted against session number (Figure 38). Negative values of *AI* indicated stronger responses to the sample than to the test, while positive values indicated the opposite.

A Spearman's rank correlation between *AI* and session number revealed that over the course of training, the *AI* became less negative for monkey 1, at the V4 location, whereas it became more negative for monkey 2, at the V1 location (Table 14). In other words, for monkey 1 at the V4 location, training was accompanied by

significant decreases in the degree of adaptation, for all three test contrasts, whereas for monkey 2 at the V1 location, effects of adaptation became stronger.

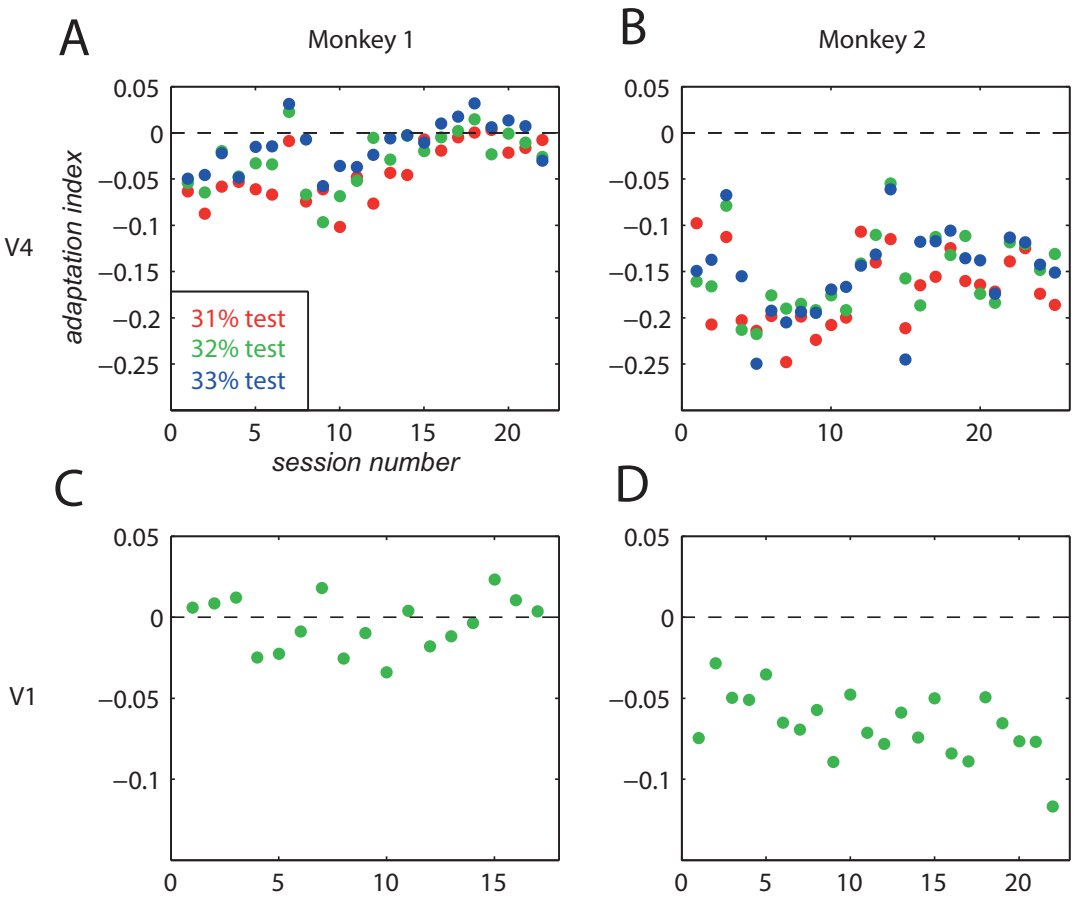


Figure 38. Adaptation indices as a function of session number, revealing changes in contrast adaptation over the course of training in monkey 1 (A & C) and monkey 2 (B & D). A & B: V4 location; C & D: V1 location. *AI*s of less than 0 correspond to weaker test-induced than sample-induced activity; *AI*s above 0 correspond to the opposite.

	Test contrast (%)	Correlation					
		df	*r*	*q*	*df*	*r*	*q*
		Monkey 1			Monkey 2		
V4	31	20	.714	< .001*	23	.248	.230
	32	20	.537	.0110*	23	.395	.0514
	33	20	.584	.00502*	23	.316	.124
V1	32	15	.0515	.846	20	-.492	.0214*

* $q < \alpha$

Table 14. A Spearman's rank correlation analysis was calculated to assess whether the differences in firing rate to sample and test stimuli changed with time. For monkey 1,

when stimuli were presented at the V4 location, adaptation effects decreased with training for the sample contrast conditions of 31 and 32%, whereas they increased for monkey 2, when stimuli were presented at the V1 location (FDR correction, $\alpha = .05/8 \times 4 = .025$).

In summary, adaptation in monkey 1 was present for some V4 channels, but signs of facilitation were also found in a number of channels. Across the population, adaptation effects diminished over time. Responses from V1 channels in monkey 1 were largely indistinguishable between those evoked by the test and those evoked by the sample. In monkey 2, on the other hand, adaptation was found in both V4 and V1, and effects of adaption in V1 increased with time.

1.6.11 Response adaptation prior to stimulus onset

1.6.11.1 Pre-stimulus adaptation

The previous section examined whether stimulus-evoked activity showed signs of adaptation, and whether levels of adaptation changed during learning. A further investigation was carried out to assess whether changes in *pre*-stimulus spontaneous activity levels changed over the course of training.

Results differed substantially between the two subjects. For monkey 1, when stimuli were presented at the V4 location, all of the channels displayed higher activity during the pre-test period than during the pre-sample period (29/29 channels). When stimuli were presented at the V1 location, most of the channels showed higher responses during the pre-test period, compared to the pre-sample period (17/23 channels showed significant differences in pre-sample and pre-test activity [3 with higher responses during the pre-sample than to the pre-test period, and 14 with higher responses during the pre-test period]).

For monkey 2, however, when stimuli were presented at the V4 location, the majority of channels displayed lower activity during the pre-test than during the pre-sample period (19/20 channels [18 with higher responses during the pre-sample than to the pre-test period, and 1 with higher responses during the pre-test]). Similarly, when stimuli were presented at the V1 location, most of the channels showed lower responses

during the pre-test, compared to the pre-sample, period (24/25 channels [all 24 had lower pre-test responses]).

When data were combined across channels, the results obtained at the population level at both locations matched those seen for individual channels (Figure 39). For monkey 1, responses during the pre-test were substantially higher than those seen during the pre-sample period (V4: $t(606216) = 418.6$, $p < .001$; V1: $t(310339) = 24.8$, $p < .001$), whereas the opposite pattern was observed for monkey 2- responses were lower during the pre-test than during the pre-sample period, indicating that activity was suppressed prior to test stimulus onset (V4: $t(400820) = -101.1$, $p < .001$; V1: $t(552450) = -136.5$, $p < .001$).

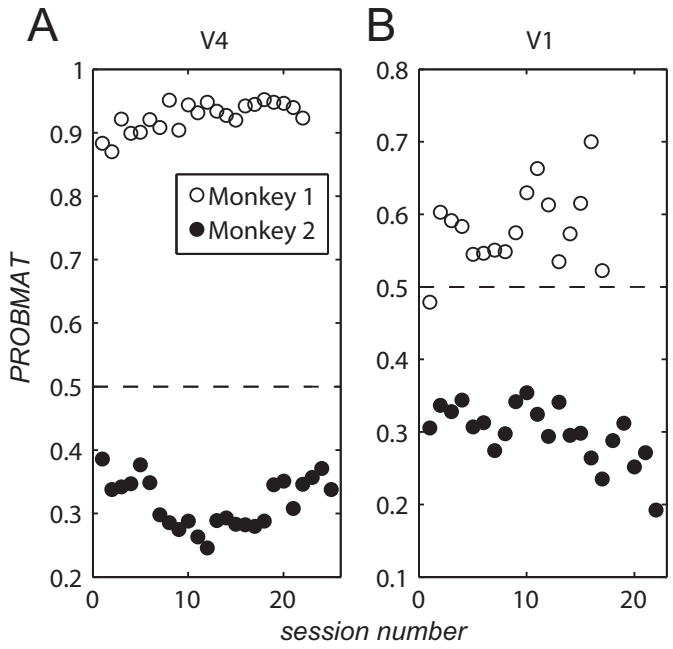


Figure 39. PROBMAT values (based on population activity combined across channels), comparing pre-sample with pre-test activity, as a function of time. A PROBMAT value of 0.5 indicates that the levels of activity during the pre-sample and pre-test periods were identical. Values above 0.5 correspond to higher pre-test than pre-sample activity, while values below 0.5 indicate the opposite. When stimuli were presented at the V4 location (A), in monkey 1 (unfilled markers), PROBMAT values started at relatively high levels (around 0.88), and increased even further as training progressed, indicating that firing rates during the inter-stimulus-interval grew stronger, relative to pre-sample firing rates. No changes were observed at the V1 location (B), where PROBMAT values were scattered at around 0.6 throughout training. For monkey 2 (filled markers), PROBMAT values were below 0.5 at both recording locations, and were further reduced with training at the V1 location.

1.6.11.2 Changes in pre-stimulus adaptation with training

A Spearman's rank correlation analysis was performed between PROBMAT values and session number, to assess whether the differences between pre-sample and pre-test firing rates changed with time. For monkey 1, when stimuli were presented at the V4 location, PROBMAT values were found to increase with training, indicating that activity levels rose during the inter-stimulus-interval (ISI) (V4 location: $r(20) = .651$, $p = .00135$; V1 location: $r(15) = .24$, $p = .352$), whereas for monkey 2, when stimuli were presented at the V1 location, PROBMAT values decreased further with training, indicating that activity levels during the ISI became more strongly suppressed over time (V4 location: $r(23) = .002$, $p = .996$; V1 location: $r(20) = .635$, $p = .00189$). All alpha levels were FDR corrected for multiple comparisons, $\alpha = .05/4 \times 2 = .025$.

In summary, responses were elevated during the sample-test interval, compared to during the spontaneous activity period prior to sample onset, in monkey 1- regardless of the location of the stimuli. A different pattern emerged in monkey 2- activity during the sample-test interval was reduced, compared to that seen prior to sample presentation.

1.6.12 Test-test discriminability

Thus far, the analyses were carried out based on a comparison of activity evoked by sample and test stimuli, between the two stimulus presentation intervals on each trial. In theory, the task could have been designed with only one stimulus presentation interval, such that only the test stimulus appeared, forcing the subjects to learn by trial and error, and build up an internal reference of 30% contrast, based solely on the feedback generated by the reward delivery system. In practise, this was not the method chosen, as it would have been much more challenging and time-consuming for the subjects to generate and make comparisons with such a template, than to be presented with real, physical stimuli. However, in order to explore the notion that PL may have been accompanied by enhanced spike discriminability during the test presentation period alone, an examination of spike discriminability was carried out by comparing test-evoked responses to stimuli of contrasts flanking 30%. This necessarily involved a

pooling of data across trials, thus AUROC values were used as the measure of spiking discriminability, for this section.

For the V4 data, activity levels that were elicited by test stimuli of 29% contrast were compared to those of 31% contrast. For the V1 data, responses were compared between test stimuli of 28 and 32% contrast. To monitor levels of test-evoked spike discriminability over the course of training, AUROC values were computed for each channel and session, and plotted against session number, for individual data. This procedure was carried out separately for trials with correct and incorrect responses.

A Spearman's correlation was performed to identify channels in which test-test discriminability changed over time. In monkey 1, 1/29 V4 channels showed a significant increase in AUROC for correct trials, and a simultaneous decrease for incorrect trials, while 3/29 channels showed significant decreases for incorrect trials. In monkey 2, 6/20 channels showed significant increases for correct trials. At the V1 location, in monkey 1, none of the V4 channels showed significant increases in AUROC during correct trials, but 2/23 channels showed significant decreases during incorrect trials. For monkey 2, 11/25 channels showed significant increases for correct trials, and 5/25 channels showed significant decreases for incorrect trials. In summary, all significant changes took place in the predicted direction (increases in AUROC for correct trials and decreases in AUROC for incorrect trials).

Figure 40 depicts the changes observed on two example channels from monkey 2 (one from V4 and one from V1). Over the course of training, AUROC values diverged significantly from 0.5 for correct trials, indicating that for these example channels, PL was accompanied by greater spike discriminability to the two test contrasts being compared.

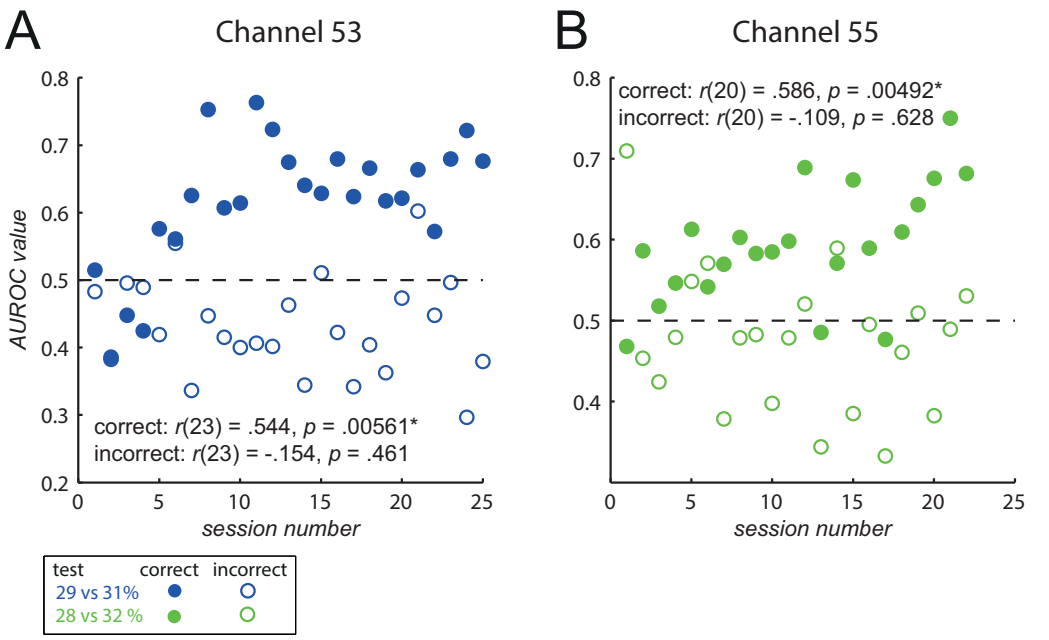


Figure 40. AUROC values, comparing test-evoked activity, for pairs of test contrast conditions: 28% versus 32% (green) and 29% versus 31% (blue). Depicted are data from two example channels, channels 53 and 55, from the V4 (A) and V1 (B) recording sites respectively, in monkey 2. Filled markers: correct trials; unfilled markers: incorrect trials.

Next, for an examination of population data, firing rates were summed across channels prior to the calculation of AUROC values, and plotted as a function of session number (Figure 41). A Spearman's correlation revealed that significant increases occurred over the course of training, at the V4 location for both monkeys, as well as at the V1 location for monkey 2, for trials in which the subject made a correct response. A significant decrease in discriminability, on the other hand, was seen during incorrect trials, for monkey 2 at the V1 location (FDR correction for multiple comparisons, $\alpha = .05/8\times4 = .025$). As with the results from the individual channel data, when significant changes occurred, AUROC values shifted away from 0.5 over the course of learning, in the directions expected.

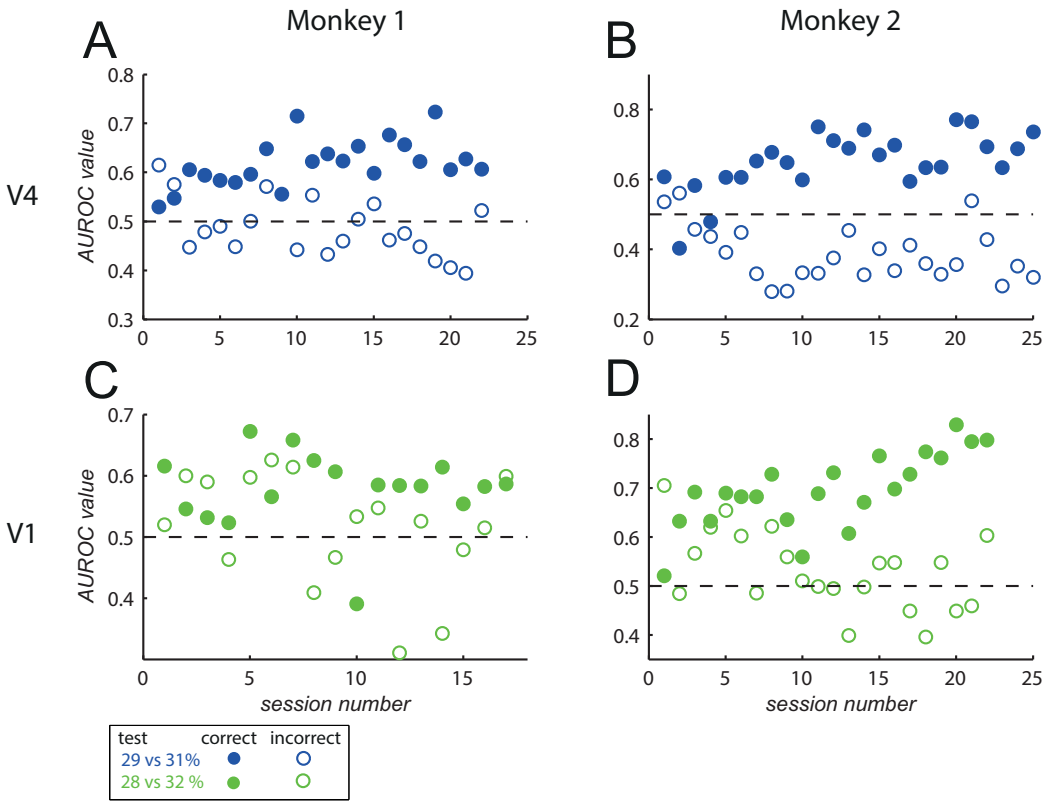


Figure 41. AUROC values for population data, comparing test-evoked activity, for 28% versus 32% (green) and 29% versus 31% (blue) test contrast conditions. Upper row: V4; lower row: V1. Left column: monkey 1; right column: monkey 2. Filled markers: correct trials; unfilled markers: incorrect trials.

	Spearman's correlation					
	Monkey 1			Monkey 2		
	r	*df*	*p*	*r*	*df*	*p*
	V4					
Correct	0.591	20	.00444*	0.582	23	.00270*
Incorrect	-0.439	20	.0424	-0.339	23	.0977
	V1					
Correct	0.027	15	.921	0.740	20	< .001*
Incorrect	-0.297	15	.247	-0.484	20	.0238*

*q < α

Table 15. A Spearman's correlation was carried out to test for changes in population test-evoked spiking discriminability over the training period, between contrast levels that flanked the value of 30% (29% versus 31% in V4; 28% versus 32% in V1).

1.6.13 Variability of the visual response

The analyses carried out thus far focused on changes in the firing rates over the course of training. To examine the possibility that behavioural improvements were accompanied not only by changes in the absolute levels of neuronal activity, but also by changes in variability of the spike response, the trial-to-trial variability of stimulus-induced spiking activity was monitored across sessions.

A two-factor ANOVA was carried out, with training period (early or late sessions) and test contrast as factors (refer to the methods section 'Test-test discriminability,' page 50, for details). When data were analysed separately for each channel, a number of channels displayed a significant main effect of training (monkey 1, V4 location: 18/29 channels [5/18 decreases, 13/18 increases], V1 location: 15/23 channels [11/15 decreases, 4/15 increases]; monkey 2, V4 location: 10/20 channels [9/10 decreases, 1/10 increase], V1 location: 10/25 channels [6/10 decreases, 4/10 increases]).

When data were combined across channels, the results matched those seen for individual channel data. Significant increases were observed over the course of training for monkey 1 at the V4 location and for monkey 2 at the V1 location, whereas significant decreases occurred for monkey 2 at the V4 location and for monkey 1 at the V1 location (two-factor ANOVA, Table 16 and Figure 42). The changes in population *FF*s for both subjects thus largely matched those seen at the individual channel level, but they did not reveal a consistent pattern between areas.

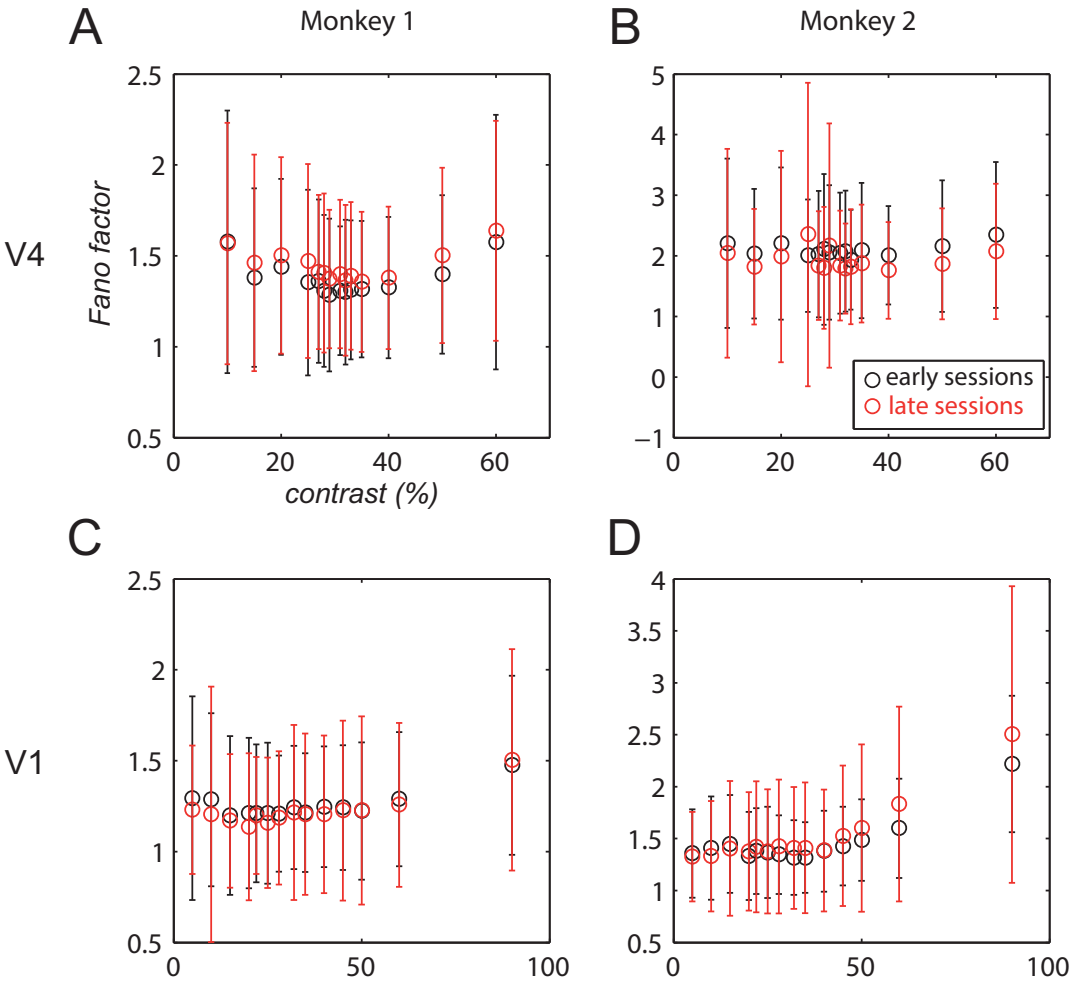


Figure 42. Population variability in spiking activity, represented by the mean Fano factor across channels, is plotted against test stimulus contrast. The *FF* increased significantly from early (black) to late (red) sessions, for monkey 1 at the V4 location (A), and for monkey 2 at the V1 location (D), whereas it decreased over the course of training for monkey 2 at the V4 location (B) and for monkey 1 at the V1 location (C, see Table 16).

		df	*F*	*q*
V4	Subject 1	1, 4844	25.9	< .001*
	Subject 2	1, 3892	17.1	< .001*
V1	Subject 1	1, 3192	4.1	.0433*
	Subject 2	1, 4172	13.2	< .001*

* $q < \alpha$

Table 16. Results from a two-factor ANOVA, comparing trial-wise spike variability between early and late sessions. The Fano factor was found to differ significantly between the two training periods, for both subjects in both locations (FDR correction for multiple comparisons, $\alpha = .05/4 \times 4 = .05$).

Note that while the response profiles obtained on individual channels appeared to consistently originate from the same subset of neurons over the course of training (as described in the section, 'Cross correlations between PSTH waveforms of channels,' on page 227), we could not conclusively verify whether the activity on each channel came solely from a single unit, or from multiple units. Thus, changes in the *FF* may have reflected not only changes in the variance of single unit spiking activity, but may also have been due to changes in the variance between responses that originated from several neurons.

1.6.14 Choice probability

1.6.14.1 Choice probability pooled across sessions

Choice probabilities were calculated for each channel (based on test-evoked activity) and plotted against time, to determine whether training would modulate the degree to which the monkeys' upcoming decision was reflected in the neuronal responses (as described in the methods section, 'Choice probability,' page 51). Calculations of CP required a sufficient number of incorrect as well as correct trials; hence this analysis focused on data obtained from the six most demanding test contrast conditions in V4 and V1 (Figure 43).

CPs closer to zero (relative to 0.5) were associated with the selection of the 'lower test contrast' target, while CPs closer to one corresponded to the selection of the 'higher test contrast' target. If the activity in a given area was indicative of the animal's upcoming choice, then one would expect CP values to be lower than 0.5 for test contrasts of less than 30%, and higher than 0.5 for contrasts over 30%. If neuronal activity in our target areas became more effective in influencing the animal's upcoming decision, then CP values for test contrasts of less than 30% should become smaller over the course of training, while CP values for test contrasts of more than 30% should become larger with time.

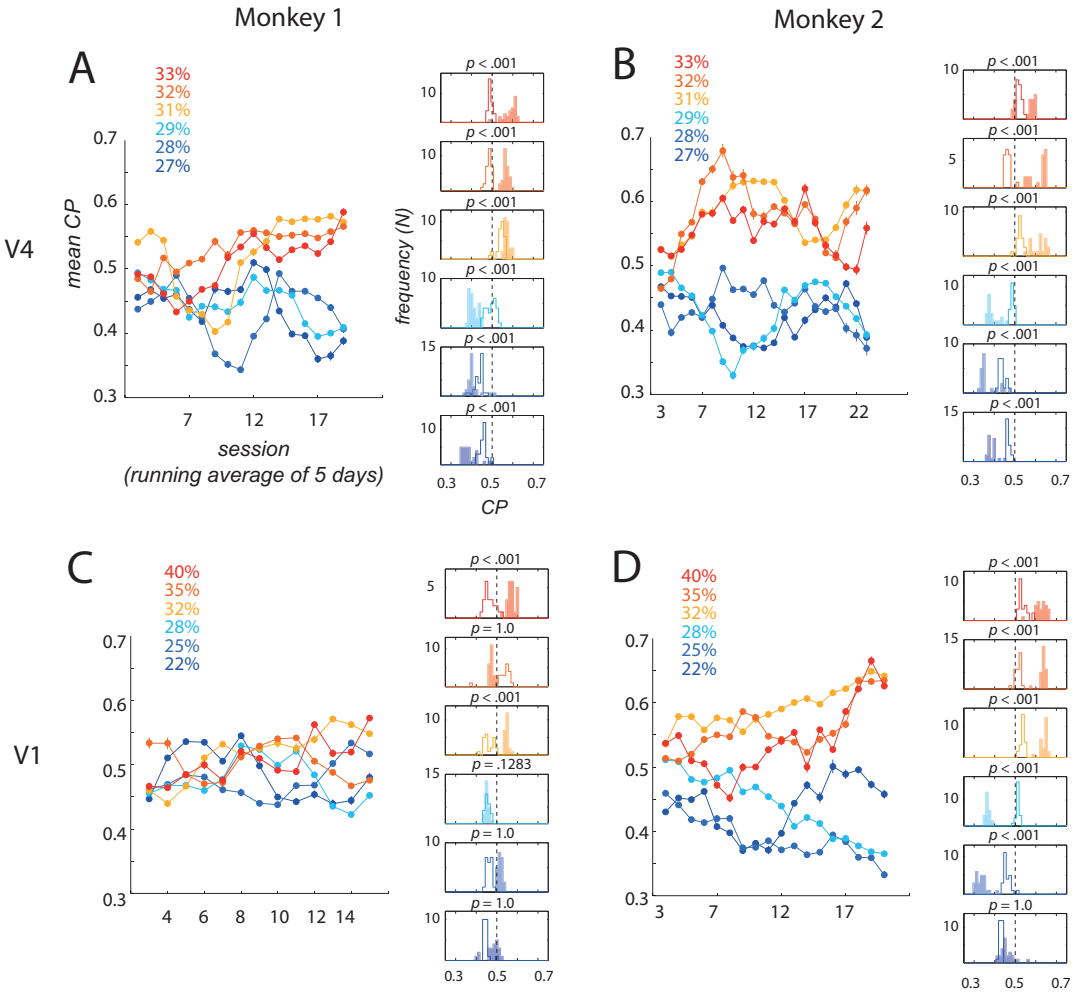


Figure 43. Main plots show CP against session number, for the hardest test contrast conditions (V4: 27, 28, 29, 31, 32, and 33%; V1: 22, 25, 28, 32, 35, and 40%; data points are colour coded according to contrast). CPs were averaged across five consecutive recording days for each channel, thus the first data point starts on day 3. Error bars show 1 SEM (note that error bars are sometimes smaller than the symbol, and thus become invisible). Small subplots (to the right of main plots) show distributions of CPs (combined across all recording channels) for a particular contrast condition. Unfilled histograms show CPs that were averaged over the first five recording days; filled histograms show CPs that were averaged over the last five recording days. *P*-values indicate whether the means of two distributions were significantly different (one-sided *t*-test).

A visual inspection indicated that neuronal activity in V4 became more indicative of the upcoming choice for all six test contrasts in both monkeys, after the initial sessions. CPs increased for test contrasts above 30%, and decreased for those below 30%. In monkey 2, after the initial divergence of CP values away from 0.5, CPs appeared to gravitate slightly towards 0.5, but data points for the higher and lower sets

of conditions remained further apart than they were at the beginning. Results for area V1 were less consistent between the two monkeys. In monkey 2, the pattern in V1 was similar to that seen in V4. For monkey 1, however, changes in CP in V1 did not consistently match those predicted. To determine whether training significantly affected the CP distributions, CPs were pooled across the first and last 5 sessions for each recording channel, for each monkey and area, and a two-way ANOVA was calculated, with training period (early or late sessions) and test contrast as factors. For both areas and both monkeys, a significant main effect of training was observed (monkey 1, V4: $F(1,336) = 775.1$, $q < .001$, V1: $F(1,252) = 11.3$, $q < .001$; monkey 2, V4: $F(1,228) = 606.0$, $q < .001$, V1: $F(1,289) = 1911.6$, $q < .001$, FDR correction for multiple comparisons, $\alpha = .05/4 \times 4 = .05$).

To investigate whether neuronal activity became more or less indicative of the upcoming choice, and whether this depended on the test contrast (i.e. on the difficulty of the discrimination required), post-hoc one-sided *t*-tests were performed to compare the means of the distributions between early and late sessions. These distributions are shown in the small subplots of Figure 43, along with the associated *p*-values. Without exception (i.e. for all six test contrast conditions), CPs in V4 became more informative of the upcoming choice in both monkeys. This was also the case for most of the test contrasts in V1 for monkey 2. For area V1 in monkey 1, however, no consistent pattern in the direction of shift of CPs was observed.

In summary, with training, CP values in V4 became increasingly representative of the animals' upcoming choice, and the magnitude of changes observed could be as large as 0.1, i.e. when an ideal observer used single-trial activity to predict the animal's choice, the observer's performance increased by around 10% over the course of training. In V1, however, CP values did not become more indicative of the animal's upcoming choice across all test contrasts in monkey 1, although this was the case to a large extent in monkey 2.

1.6.15 Control analysis conducted to assess declines in response discriminability with time

1.6.15.1 Stability of responses to oriented stimuli over the course of training

The overall decrease in CRF maxima observed on several channels (as reported in the section, '*Changes in the CRF for individual channels*,' page 53) may potentially have been caused by a general decline in the quality of the recording signal (Rousche & Normann, 1998), due to changes such as biological encapsulation of electrodes (Anderson, 2001) and mechanical injury to cortical tissue (Polikov, Tresco, & Reichert, 2005; Rousche & Normann, 1998).

To determine whether this was the case, it was necessary to compare the quality of neuronal responses across sessions from a task that did not involve perceptual learning, and which was independent of the CD task. Thus, responses to grating stimuli that were presented during the passive viewing task for the mapping of spatial frequency and orientation tuning preferences in our channels were monitored throughout the training period.

1.6.15.2 Methods

For these signals, spike thresholds were set manually using CSC Spike Extractor software (Neuralynx, Inc.), and levels of spontaneous activity were not intentionally constrained to fall within a predefined range across sessions. This precaution was taken to reduce the likelihood of introducing artificial similarities in the data across sessions. Our aim was to obtain spike signals that had undergone as little processing as possible; although the manual method of sorting spikes introduced some variability in spike processing from day to day, the underlying assumption was that this variability was unlikely to follow a systematic pattern over the training period.

The preferred SF, phase, and orientation were obtained for each recording session, and data were compared across training sessions, allowing us to examine whether activity levels on each channel remained consistent throughout the months of training, during the passive viewing task. To normalise responses against baseline firing

rates, levels of spontaneous activity were subtracted from stimulus-evoked spiking activity for each channel.

1.6.15.3 Results

Mean responses to stimuli of the PO were compared across sessions, and the combination of SF and phase that elicited the highest mean firing rate was identified. Spiking activity elicited by this particular combination of stimulus properties (PO, and optimal SF and phase, given the PO) was plotted against session number, and a Spearman's rank correlation was calculated to identify changes in activity with time. A few channels showed significant changes in each area (Table 17); in total, 6/11 channels showed significant increases in activity, while 5/11 channels showed significant decreases.

			Spearman's rank correlation		
Area	Channel #	*PO*	*df*	*r*	*p*
V4			Monkey 1		
	52	0	20	-.49	.0160
			Monkey 2		
	5	30	15	-.51	.0148
	10	25	15	-.49	.0205
V1			Monkey 1		
	8	120	23	.78	< .001
	9	73	23	-.50	.0493
			Monkey 2		
	7	68	20	.78	< .001
	14	28	20	.55	.0135
	18	15	20	.47	.0391
	22	45	20	-.63	.00339
	26	59	20	.55	.0129
	51	91	20	.49	.0292

Table 17. List of channels for which levels of spiking activity in response to stimuli presented during a passive viewing task underwent significant changes over the training period.

In conclusion, changes in activity during a passive viewing task were seen on a small minority of channels, and when they did occur, activity levels tended to increase with time for V1 channels. This pattern indicated that the amplitude of the recording

signal remained high throughout the training period, and showed few signs of decline with time.

1.6.16 Discussion of neuronal results from the CD task

Our results showed clear distinctions in the changes that occurred at the neuronal level, between areas V4 and V1. Changes at the neuronal level were more prominent in monkey 2 than in monkey 1- a finding which agreed well with the behavioural performance of the two subjects (refer to the section, '*Perceptual learning for individual test contrast conditions*,' page 28), in which gains occurred over a longer period of training in monkey 2 than in monkey 1.

In V4, we observed a steepening of contrast response functions and shifts of the C_{50} towards 30% in monkey 2, along with shifts of the PNE towards 30% in both monkeys. These changes corresponded to improvements in contrast sensitivity and spiking discriminability around the most difficult contrast levels used during training (i.e. those closest to the sample contrast). Our results were reminiscent of changes in activity of orientation-selective V4 neurons that have been reported by previous studies, during training on orientation discrimination tasks; namely, a sharpening of tuning curves (Yang and Maunsell, 2004), and an increase in signal discriminability (Zivari Adab and Vogels, 2011). The improvements in V4 spike discriminability that we observed in our animals as training progressed thus closely matched their enhancements in CD ability.

Across the population of channels, little change was seen in maximum activity levels in either location, whereas for individual channel activity, differing effects were observed for the two subjects. At the V4 location in monkey 1, maximum responses tended to decrease, while in monkey 2, they tended to increase. Thus, our findings differed somewhat from the increases in net firing rate seen by Raiguel (2006) in V4 during a fine orientation discrimination task (though Zivari Adab and Vogels (2011) reported mild reductions in overall response strength during training with coarse orientation discriminations).

The improvements in the CRF and the PROBMAT function that we observed in V4 were similar to those seen in the CRFs of V1 neurons in anaesthetised cats by Hua et al. (2010). However, in our study of macaque V1, these effects did not occur consistently; rather, we found that the pattern of change depended on the subject. In monkey 1, few changes were observed at either the population or the individual channel level; in monkey 2, while the slope of the CRF increased with training, the C_{50} tended to shift away from 30%. Both subjects showed significant improvement on the CD task at the V1 as well as at the V4 location, thus the lack of a shift in the C_{50} towards the sample contrast in monkey 2 did not appear to be linked to poor task performance.

Furthermore, in V1, changes in absolute firing rates for individual channels were seen predominantly for monkey 2, but not for monkey 1. For monkey 2, the pattern of declining activity observed in V1- reflected by a decrease in the maximum of the CRF in a number of individual channel recordings- were reminiscent of those seen for V1 neurons in previous studies on perceptual learning during an orientation discrimination task (Ghose et al., 2002; Schoups et al., 2001). (In our study, note that spontaneous activity levels were standardised across training days, during the spike threshold selection process. Thus, we were not able to determine whether these changes were best described by a contrast gain, a response gain, or an additive model- refer to the section, '*Automated threshold setting to obtain uniform spontaneous activity levels* across sessions,' page 218, for details.)

Our results thus support models of learning in which widespread, task-related changes that are able to account for improvements in perceptual ability occur predominantly in intermediate areas such as V4. They also show that accompanying changes take place in V1- on a lower rung of the visual hierarchy- but these changes do not appear to be as directly linked to behavioural improvements as those seen in V4. Thus, multiple areas are affected during perceptual learning of a contrast discrimination task, and the contributions made by V4 neurons differ from those made by V1 neurons.

In addition to the changes seen in the CRF and in stimulus discriminability, task training was characterised by changes in response adaptation (less adaptation was seen over time in V4 for monkey 1, whereas more adaptation was seen in V1 for monkey 2); these differences may indicate an adoption of differences in task strategies by the two

subjects. In monkey 2, for example, test-evoked responses diminished with learning in V1; this seems to show that the dynamic range of activity to different test contrasts was reduced. However, we also observed decreases in spontaneous activity levels during the pre-test period, which occurred in tandem with increases in stimulus adaptation. Thus, if monkey 2 had learnt to make an evaluation of test contrast based on the *difference* between test-evoked responses and pre-test levels of spontaneous activity, rather than being based solely on a comparison of absolute levels of test-evoked activity, then response adaptation would have enhanced his decoding abilities.

We also found changes in the levels of noise correlation between simultaneously recorded channels, with training. Noise correlations are often assumed to hinder the decoding of sensory information, as neuronal responses vary substantially from trial to trial. If neuronal responses are uncorrelated between channels, this variability could effectively be reduced through an averaging of responses across neurons- provided that a sufficient number of neurons are pooled together (Cohen & Kohn, 2011; Cohen & Maunsell, 2009; Gu et al., 2011; Mitchell, Sundberg, & Reynolds, 2009; Shadlen & Newsome, 1996). Thus, a reduction in noise correlation with training might lead to higher accuracy in the decoding of signals, which might in turn improve the animal's performance.

In V4, we found a decrease in the level of noise correlation with training, in monkey 2 (note that when the period was training was divided into two halves, a significant reduction was also obtained in monkey 1). These findings are in line with reports from Gu et al. (2011), who found that noise correlations are lower in area MST of trained animals, compared to untrained animals. However, the authors also report that the reductions were unable to fully account for the improvements observed at the behavioural level. We have not performed the equivalent decoding analysis, so cannot determine whether this was also the case in our data.

Surprisingly, we found that noise correlations increased significantly in V1 in both monkeys. This increase in noise correlations would presumably impair the decoding of population activity if high levels of correlation are present in the signal, although if low levels of correlation are present, it might even be beneficial (Averbeck, Latham, Pouget, 2006). We found that the location of the half-maximum point of

contrast tuning curves in V1 became more variable with training, in monkey 2. A more in-depth analysis of the amount of signal correlation over the course of training, and its possible relationship with changes in noise correlations, is needed to shed additional light on these questions.

Training was also accompanied by increases in choice probability in V4 for both monkeys, and in V1 for monkey 2. Correlations between levels of neuronal activity and a subject's upcoming behavioural choice have been observed across in a range of cortical areas (for a review, see Nienborg, R. Cohen, and Cumming (2012)). One might expect the strength of such modulations to be greater at higher-order cortical regions; for instance, in a task involving perceptually ambiguous stimuli, Grunewald et al. (2002) observed that the proportion of V1 neurons with behavioural-choice-dependent modulations was around a third of that seen in MT. Palmer, Cheng, and Seidemann (2007) reported CP values in V1 MUA of around 0.62, when macaque subjects performed a reaction-time visual detection task, while Shiozaki, Tanabe, Doi, and Fujita (2012) reported grand CP values of around 0.55 in V4, when subjects performed a depth disparity task.

Although our task paradigm was not specifically designed to nurture a strong mental association between the target stimuli and the test contrast, the relationship between the CW-CCW positions of our targets and the difference in contrast between test and sample stimuli remained fixed throughout training, hence the potential for learning-induced enhancements in CP modulation over time. Our results confirm those of previous studies which found that neuronal signals at both low- and intermediate-level regions of the visual hierarchy were able to represent upcoming decisions about behavioural choices, in addition to being able to encode stimulus-related information. This leads to the question of the extent to which changes in CP were attributable to general task learning (Uka, Sasaki, & Kumano, 2012), and the extent to which they were direct accompaniments of fine perceptual learning. Task learning would be expected to occur predominantly during early sessions, i.e. during initial training at the V4 location, and to eventually decline to zero for subsequent sessions. The fact that changes in CP were significant in monkey 2 during training at the V1 location indicates that the increase in CP values was not merely due to procedural and associational learning, but could also occur during learning of fine contrast discriminations.

We further demonstrated that the strength of these representations can be enhanced through PL in a test-contrast-dependent fashion, i.e. the greater the difference between the test and sample contrasts, the larger the increase in CP. In some cases, we observed changes in CP from around 0.5 at the start of training, to around 0.65, in both V4 and V1. Similar changes in CP as a result of training on a perceptual task have previously been documented in MT (Dodd, Krug, Cumming, & Parker, 2001), although not in V4 (Zivari Adab & Vogels, 2011).

1.6.16.1 Techniques used for the analysis of spiking activity

In the current study, several methods of analysing spiking activity were compared, to maximise the extraction of contrast-dependent information and to optimise levels of discriminability between stimulus-evoked responses. Levels of spiking activity between channels, as well as within individual trials, were positively correlated, thus the extraction of information benefitted from a pooling of responses across channels. Noise correlations in V1 increased over the course of training in both animals; in V4, on the other hand, they tended to decrease (the effect in V4 was also more prominent in monkey 2 than in monkey 1).

Our PROBMAT method was robust to between-trial fluctuations in activity and exploited correlations in firing rate across channels. Furthermore, it adopted a biologically-realistic approach, as it reflected the fact that the animals had access to information from a large population of neurons at any given time, rather than to the activity of a single 'highly-performing' neuron, pooled across many repetitions of trials.

We also examined the outcomes of different approaches in the normalisation of spiking activity. In electrophysiological studies such as this, when population firing rates are calculated, activity is often normalised to the maximum responses of individual neurons, prior to being combined across channels, in order to compensate for potential biases in data recording, e.g. in the orientation and proximity of electrodes to cell bodies, and sampling biases. Overall response strengths are dependent on two variables: 1) the inherent level of responsiveness of a neuron, and 2) the proximity of the neuron to the sampling electrode. These two factors could potentially interact to influence the discriminability of recorded spiking activity. For example, if the neurons closest to our

electrodes happened to have steep CRFs and neurometric functions around the contrast of 30%, while distally-located cells had shallower CRFs and neurometric functions, this would yield high levels of discriminability in the signal. If, instead, the neurons with shallower slopes were located close to our electrodes, while those with steeper slopes were further away, then the level of discriminability in our signal would be poorer.

Although we could not address this issue directly by adjusting the position of our electrodes, we carried out a comparison of spike discriminability before and after normalisation, and found that discriminability degenerated after a process of normalisation, for V4 data in one monkey, and for V1 data in the other monkey. This may have reflected inherent differences in the response strengths of individual neurons, e.g. the most informative neurons may also have been those with the highest firing rates. Top-down attention is known to increase firing rates of contrast-responsive neurons, creating a similar effect to that achieved by an increase in stimulus contrast (as described in the literature review for this chapter in the section titled, '*Effects of attention on contrast response functions of visually-responsive neurons*,' page 12). It is thus plausible that the neurons which are most relevant to the task are also those which undergo the greatest modulations in activity.

1.6.16.2 Order of training at the V4 and V1 locations

Several studies with naïve human observers reported that improvements in contrast sensitivity were specific to the retinal location used during training (Sowden et al., 2002; Xiao et al., 2008; Yu et al., 2004). In addition to replicating the finding that PL in a CD task is location-specific, Xiao et al. (2008) carried out a 'double training' paradigm in which they showed that significant transfer of CD learning to a new retinal location was possible, if the new location had been 'primed' by prior training on an orientation discrimination task. The researchers suggested that this transfer may have arisen from enhanced deployment of spatial attention at the new location. Subsequent research from Yu's lab has demonstrated this transfer of learning in an orientation discrimination task and a Vernier task (Wang, Zhang, Klein, Levi, & Yu, 2012; T. Zhang et al., 2010).

In our study, presentations of grating stimuli at the V1 location prior to V1 training were limited to passive viewing conditions, with fixation at centre, during a spatial frequency mapping paradigm, as reported in the section, '*Control analysis conducted to assess declines in response discriminability with time*,' page 107). Hence, subjects had not performed any tasks at the V1 location, prior to training with stimuli at this location. Based on the results from the human studies, it seemed unlikely that our subjects' performance at the V1 location would benefit significantly from the practise undertaken at the V4 location; however, several caveats should be noted.

In our task, training at the V4 location spanned numerous sessions (monkey 1: 30 sessions; monkey 2: 25 sessions), unlike that carried out the human studies, which spanned two sessions for the task carried out by Yu et al. (2004), five to six sessions for Xiao et al. (2008), and ten sessions for Sowden et al. (2002). (Tsodyks et al. (2004) reported that one of their subjects practised a CD task for 40 sessions, without any improvement in CD thresholds, though this might have been an exceptional case.) Furthermore, the fact that training-induced transfer of learning was not visible at the behavioural level does not guarantee absence of change at the neuronal level. It is conceivable that changes in V1 might have undergone modulations as a result of training at the V4 location; in our task, it was not possible to test this due to the non-overlapping RFs of our V4 and V1 neurons.

1.6.16.3 Stability of signals over the training period

A limitation of using chronically implanted arrays is the inability to adjust electrode position and depth, to optimise signal quality. A central goal of our study was to monitor learning-induced changes in activity over time, thus any deterioration of the signal due to biological reactions or mechanical failure was a concern, as that would be likely to introduce systematic reductions in SNRs, which might in turn be incorrectly interpreted as decreases in discriminability during learning.

To address this issue, spike recordings were carefully examined and artifacts were removed from the data wherever possible (refer to the section '*Artifact removal from neuronal data*' on page 213 for a detailed description of how artifacts were removed). A decline in stimulus-evoked activity was seen in a number of V1 channels;

to verify that this was due to the effects of perceptual learning, rather than to an intrinsic deterioration of the signal, a passive viewing task was carried out during each session, immediately before training on the CD task commenced. This analysis found little change in stimulus-evoked activity over time, giving support to the premise that the changes in firing rate that occurred during the CD task were truly task-dependent.

The implants remained physically stable throughout the recording period; however, this was no guarantee that the multiunit activity (MUA) being sampled by each electrode remained equally consistent. Initially, to gain a very rough idea of the overall stability of the implant, the neuronal tuning properties of signals on each channel were monitored over several sessions. Receptive field locations were mapped repeatedly over several sessions in each subject, and orientation tuning remained consistent throughout the training period (‘Characterisation of neuronal tuning properties,’ page 237). However, even if the signal quality remained good throughout, the identity of the neurons from which recordings were taken might vary from one session to the next, without any discernible change in neuronal tuning preferences. To investigate this possibility in more thorough detail, a bootstrapping procedure was used to compare PSTHs from individual channels, across multiple sessions (refer to the chapter, ‘Cross correlations between PSTH waveforms of channels’ for details, page 227). The results from this analysis showed that the visually-evoked activity on a given channel was often distinctive and separable from that seen on other channels. While this method did not permit the unequivocal claim that each electrode sampled an identical set of neurons from day to day, it offered quantitative evidence that the signals remained largely consistent across recording sessions.

Chapter 2: Roving task

2.1 Roving task literature review

In a typical perceptual discrimination task, two stimuli are presented per trial, in two separate time intervals, and subjects compare the stimuli based on a property such as orientation or contrast. In many cases, one of the stimuli retains the same appearance for a prolonged period of time (such as for a block of trials, or for an entire session), whereas the other stimulus varies in the parameter of interest from one trial to the next.

However, this is not always the case- in some studies, properties of the stimuli presented during *both* intervals are allowed to vary, so that neither one remains constant across a large number of consecutive trials. This task paradigm is termed 'stimulus roving' (Berliner & Durlach, 1973; Parkosadze et al., 2008).

2.1.1 Stimulus roving during contrast discrimination tasks

Most of the previous studies on stimulus roving have demonstrated that when stimulus features vary unpredictably from trial to trial, this generally makes a task harder to learn- improvements in performance are slower, diminished, or absent altogether (Adini et al., 2004; Kuai et al., 2005; Otto, Herzog, Fahle, & Zhaoping, 2006; Parkosadze et al., 2008; Yu et al., 2004). Based on these reports, it has been hypothesized that a stimulus roving paradigm might not merely reduce the rate of learning; it might also exert an inhibitory effect on perceptual learning and thus actively impair task performance. In contrast discrimination tasks carried out by Adini et al. (2004) and Yu et al. (2004), for instance, training lasted for 4 to 5 practice sessions, and learning was severely limited under roving conditions. A possible explanation for this phenomenon (Kuai et al., 2005; Yu et al., 2004) is that memory traces are continually disrupted when the pedestal contrast (termed the 'sample contrast' in our study) varies across trials, thus preventing observers from constructing and maintaining internal reference templates to which they would otherwise refer.

Yu et al. (2004) trained two groups of naïve subjects on a CD task and found that while training on a non-roving task produced clear, significant improvements in

contrast thresholds, training on a roving task generally resulted in much less improvement, if any. Two of their four subjects showed no improvement; one showed some improvement for low contrasts; and the last showed improvement for high contrasts but a worsening of thresholds for low contrasts.

Adini et al. (2004) trained their subjects on a blocked (i.e. non-roving) multi-pedestal task, and then tested them on a roving task that was more demanding than Yu et al.'s task, with a total of seven randomly-interleaved pedestal contrasts. They observed barely-significant changes in threshold, the directions of which were dependent on the pedestal contrast. The researchers hypothesised that their subjects may have adapted their task strategy over the course of training- in the face of uncertainty regarding the pedestal contrast, subjects may have tackled a subset of pedestals contrasts at a time, e.g. by focusing on low-to-intermediate contrast discriminations during early sessions, and then tackling higher-contrast discriminations once the lower-contrast conditions had been mastered. Improvements may have reflected changes in the shape of the contrast transducer function of individual neurons; alternatively, they may have resulted from changes in connectivity *between* neurons, through an optimisation in the selection and gating of subpopulations of channels.

Subsequent work by Kuai et al. (2005) and J.-Y. Zhang et al. (2008) (both from Yu's group) introduced variations in the timing structure of stimuli used during training, and found that the temporal patterning (or lack thereof) of roving stimuli was able to significantly influence the amount of learning observed. When the reference contrasts (referred to in our study as 'sample contrasts') were varied from one trial to the next according to a fixed, recurring sequence, improvements in CD during this 'temporally patterned' task were possible; on the other hand, if the temporal sequence was modified such that inter-trial intervals of unpredictable durations were used, then learning was inhibited. Furthermore, if subjects were given an additional practise session within a few hours of undergoing training on a temporally patterned task, in which sample contrasts were randomly interleaved across trials (termed 'mixed-by-trial' or 'MBT' training), this later period of exposure to roving contrasts impaired learning-presumably due to disruption of memory consolidation (J.-Y. Zhang et al., 2008). The researchers suggested that these findings offered support for Ahissar and Hochstein's RHT model, as the regular temporal ordering of reference contrasts might facilitate the

'tagging' of stimuli and enable top-down attentional mechanisms to target low-level cortical regions during PL-induced plasticity.

2.1.2 Insights from a roving paradigm during a bisection task

Otto et al. (2006) approached the roving issue from a theoretical modelling standpoint. They predicted that perceptual learning for a bisection task would be disrupted by stimulus roving, based on the assumption that the brain has limited resources. Although subjects were capable of making rapid improvements with one outer-element-distance condition at a time, the researchers hypothesised that their perceptual processing machinery became overwhelmed when presented with multiple outer-element-distances simultaneously. The researchers suggested that conflicting inhibitory and excitatory mechanisms underlie the changes required for different pedestal conditions and end up interfering with each other, thus impeding learning.

Following this, Parkosadze et al. (2008) trained subjects intensively on a similar bisection task to that used by Otto et al. (2006), using just two pedestal conditions, and for a much longer period, in which observers completed 18,000 trials in total. When they examined the initial subset of data collected from the first 3,600 trials, they obtained results that replicated those of Otto et al. closely- in some cases, they even observed a slight deterioration of bisection acuity thresholds. With further practise, however, the subjects did eventually improve at the task, and their thresholds reached similar levels to those obtained during non-roving task training. Thus, while the brain initially seems to have difficulty when presented with roving stimuli, given enough time and practise, it may adapt to the task at hand.

In summary, it was originally thought that the learning of a roving task was near-impossible; recent studies, however, have revealed that perceptual learning is, in fact, possible under roving conditions, and that the pace of learning is influenced by the temporal structure of stimulus presentation. Extensive practice may eventually yield cumulative improvement- regardless of whether the parameter of interest remains stable across consecutive presentations (e.g. a block of trials), or fluctuates rapidly between rival conditions.

2.1.3 Goals of the roving task

The roving tasks implemented by Adini et al. (2004) and Yu et al. (2004) involved around 4 to 5 training sessions, and learning was severely limited under roving conditions. Our aim was thus to investigate the effects of a prolonged and more intensive period of roving task training in macaque subjects, spanning several weeks.

We could not explicitly instruct our monkeys to base their decisions on comparisons between the sample and test stimuli and disregard the rules learnt during prior non-roving training (e.g. 'make a comparison with a 30% sample'). This constraint further necessitated a longer period of training, in which subjects obtained feedback via reward delivery and were conditioned to perform according to task requirements.

If dramatic improvements in performance proved possible, then we reasoned that human subjects might benefit from an extension of the training period (as with the results seen by Parkosadze et al. (2008)). On the other hand, if our results were similar to those reported in the contrast domain by Adini et al. and Yu et al., then the lack of substantial improvement may be attributable to the roving paradigm itself.

As before, neuronal activity was monitored throughout training. To our knowledge, investigations of the neurophysiological changes that occur during roving training have not been carried out before, making this a new topic of research. As we intended to introduce flanker stimuli to the roving task at a later stage (see Chapter 3: Flanker task), and the CRT monitor used for stimulus presentation was not large enough to accommodate flankers for stimuli positioned at the location of the V4 RFs, training on the roving task was carried out solely with stimuli positioned within the V1 RFs.

We had already observed significant changes in spiking activity in both V4 and V1, as described in the previous section. Our interests were now to see whether the changes in V1 would continue under roving conditions, and whether they would correspond well with improvements at the behavioural level (if any). As previously discussed, changes in V4 activity were directly linked to improvements seen at the behavioural level, whereas changes in V1 activity appeared to be less closely related; our expectations that training-related effects in V1 would be able to account for behavioural improvements were thus conservative from the outset. Nonetheless, we felt

that even a modest or null result (such as that reported by Ghose et al. (2002) on orientation discrimination in V1 and V2, and by Law and Gold (2008) on discrimination of motion direction in MT) would make a valuable contribution to our understanding of how the learning of a CD task is implemented on a neuronal level.

2.1.3.1 Psychophysics/ behavioural questions

- Given sufficient training on a stimulus roving task, do macaque subjects show improvements in contrast discrimination?
- If so, are these improvements seen across the board, or only for certain sample contrasts?

2.1.3.2 Neurophysiological questions

- Do changes in spiking activity occur in area V1?
- What is the nature of these changes (e.g. alterations of firing rate, spike variance, and tuning properties)?
- Do neurometric and behavioural changes correlate with each other?

2.2 Psychophysics methods

To compare the effects of non-roving and roving tasks on perceptual learning, a roving stimulus paradigm was introduced.

2.2.1 Task paradigm

In the roving task, the contrast of the sample stimulus was not fixed at 30% as was done in the initial PL paradigm (Chapter 1), but could take on one of three values (20, 30 or 40%) on a given trial (referred to as the 'MBT' method in Adini et al. (2004)- a more challenging paradigm than the 'blocked' method). In turn, the test stimulus was presented at one of 12 possible contrast levels, the exact values of which depended on the contrast of the sample (20% sample: [5, 10, 12, 15, 18, 22, 25, 28, 35, 45, 60, 90% test]; 30% sample: [5, 10, 15, 22, 25, 28, 32, 35, 38, 45, 60, 90% test]; 40% sample: [5, 10, 15, 25, 32, 35, 38, 42, 45, 50, 60, 90% test]), yielding 36 conditions in total.

The requirements of the task remained the same as those described in the previous chapter on non-roving stimuli. Presentation of a sample stimulus was followed by that of a test stimulus, and subjects had to decide whether the contrast of the test stimulus was higher or lower than that of the sample.

For some conditions, the identity of the correct target was the same regardless of the sample contrast (e.g. when the test contrast was 5%, the subjects always had to saccade to the black target). However, for other conditions, the identity of the target varied, depending on the sample contrast. For example, when the test contrast was 25%, if the sample contrast had been 30% or 40%, then the subjects had to saccade to the black target, whereas if the sample contrast had been 20%, the subjects had to saccade to the white target (refer to Figure 44).

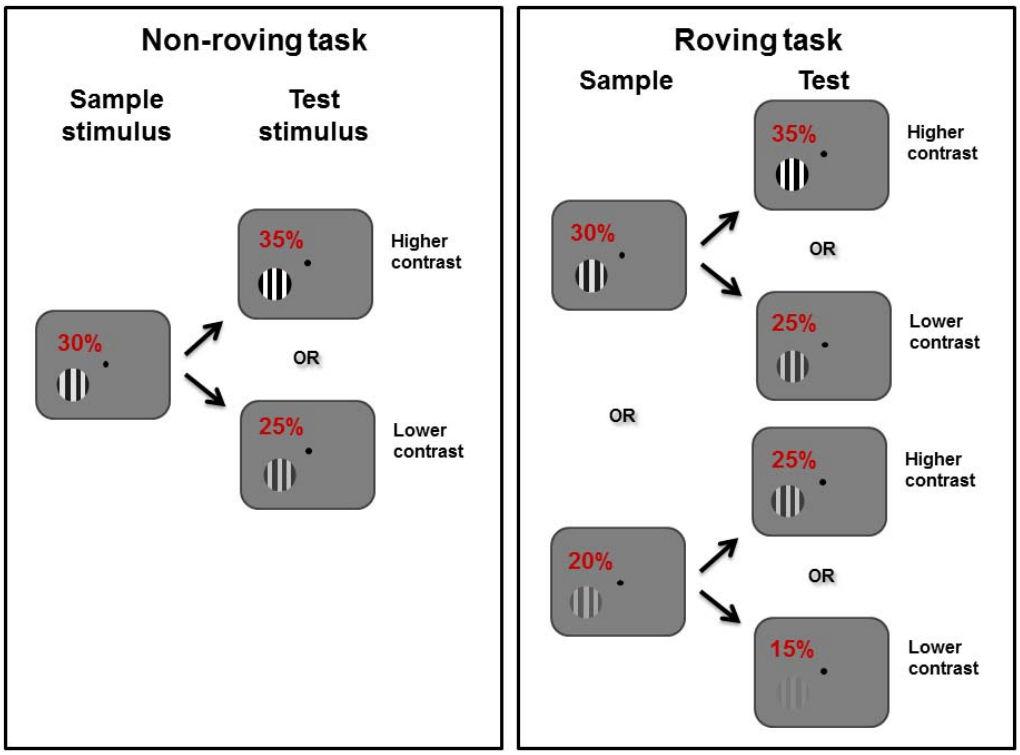


Figure 44. Characteristics of tasks involving non-roving and roving stimuli. For the task with non-roving stimuli (the 'non-roving task'), depicted in the left-hand panel, the sample stimulus always had a contrast of 30%. For the task involving roving stimuli (the 'roving task'), the contrast of the sample stimulus varied randomly from trial to trial and took on a value of 20, 30 or 40% (right-hand panel). Unlike the non-roving task, subjects had to observe the contrast of the sample stimulus in order to perform the roving task correctly. For example, for a test stimulus of 25% contrast, they were required to report that it was higher in contrast, when it had been preceded by a sample of 20% contrast, whereas they were required to report that it was lower, if the sample contrast had been at 30% or 40%.

Subjects underwent training until their performance plateaued and it seemed unlikely that additional training would bring about further improvement.

2.2.1.1 Stimuli used in the roving task

Grating stimuli were positioned in the same lower hemifield location as that used in the non-roving task, i.e. at an eccentricity of 4.6° (azimuth: -3.5°, elevation: -3°) and 1.5° (azimuth: -1.3°, elevation: -0.7°), for monkeys 1 and 2 respectively. Stimulus parameters (SF, orientation, and size) were the same as those used during the non-roving task (refer to Table 1, page 22).

2.2.2 Behavioural performance

Throughout each stage of training, the behavioural performance of the subjects was monitored to determine whether their contrast discrimination abilities had improved with training.

2.2.2.1 Measures of perceptual learning

The same measures of performance that were used in the non-roving task were applied to the data from the roving task: 1) the proportion of trials with correct responses, 2) the slope of the psychometric function, and 3) the point of subjective equality (PSE) of the psychometric function.

2.2.2.2 Reaction times

The monkeys' reaction time (RT) was monitored throughout performance of the roving task, to determine whether it was possible for this aspect of the behavioural response to undergo further enhancements over the course of training.

2.3 Neuronal methods

Methods of collecting and processing neuronal data were identical to those described in Chapter 1 (page 15).

2.4 Roving task behavioural results

2.4.1 First set of training sessions on a roving task

Subjects performed the roving task with a grating stimulus for several weeks, until their performance reached a plateau (monkey 1: 33 sessions, spanning a period of 8 weeks; monkey 2: 16 sessions, spanning 4 weeks).

2.4.2 A comparison of performance between non-roving and roving tasks, to monitor task learning

For certain conditions, the identity of the target stimulus was critically dependent on the contrast level of the sample stimulus. These conditions provided a direct means of comparing the subjects' performance before and after the introduction of the roving task paradigm. Namely, when the sample contrast was 20%, the conditions that induced this conflict were those where the test contrasts were lower than 30%, but higher than 20% (i.e. 22, 25 and 28% test contrast conditions). When the sample contrast was 40%, the conditions that induced this conflict were those where the test contrasts were higher than 30%, but lower than 40% (i.e. 32, 35 and 38% test contrast conditions). The conditions in which a conflict arose, relative to the previously learned sample contrast, are termed 'response conflict conditions.'

Note that the test contrasts of 22%, 25% and 28% were common to the 20% and 30% sample conditions, while the test contrasts of 32%, 35% and 38% were common to the 30% and 40% sample conditions. Furthermore, note that the 38% test contrast condition was introduced when roving training began, and thus no data were available for this test contrast during the non-roving period.

To aid comparison between non-roving and roving performance levels, the subjects' responses during response conflict conditions (when the sample contrast was 20% or 40%) were plotted alongside the choices made upon the presentation of 30% sample contrasts during both periods of training (Figure 45). A visual comparison revealed that at the beginning of training on the roving task, their responses to a given test contrast tended to be similar, regardless of the actual contrast of the sample, i.e.

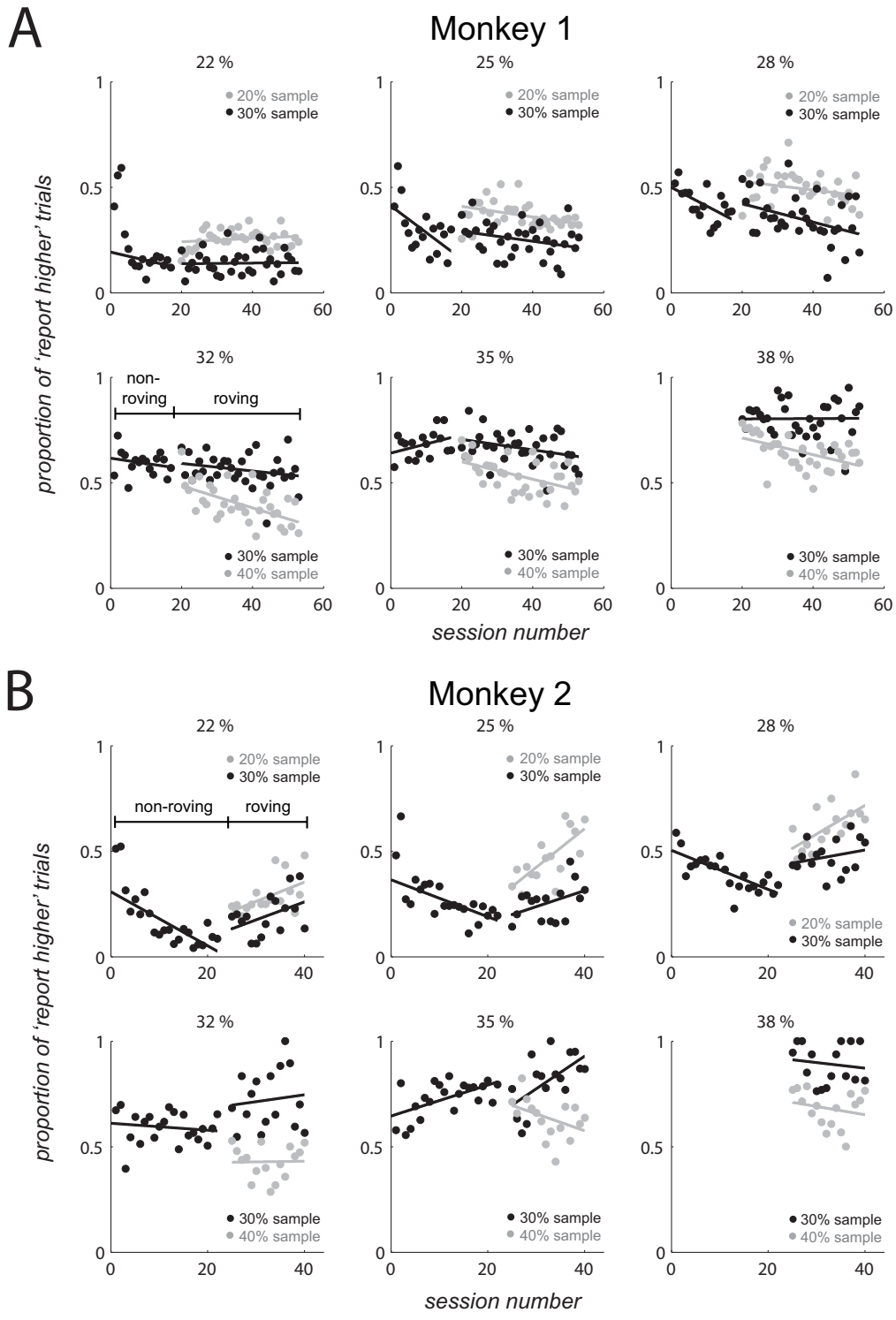


Figure 45. Proportion of trials during which the subjects reported that the test contrast was higher than the sample contrast, for conditions which gave rise to a potential conflict in responses, for monkey 1 (A) and monkey 2 (B). Within each subplot, the data points on the left indicate subjects' performance during the non-roving task, while those on the right indicate performance during the roving task. Unfilled data points:

conditions in which a 20% contrast sample was presented; medium-coloured filled data points: conditions with a 30% sample; dark-coloured filled data points: conditions with a 40% sample. A divergence in data points between response conflict conditions (represented by differences in slope between fitted lines within individual subplots) suggested that learning occurred to some degree, under roving conditions.

responses appeared to have been based on the 30% reference that was used during the non-roving task. However, the subjects' responses gradually diverged over the course of roving training, indicating that the monkeys learnt to make their comparison against the sample stimulus. Additionally, based on a visual inspection of the data, learning appeared to be more pronounced in monkey 2 than in monkey 1.

For an illustration of the patterns of change in performance, refer to the black-coloured data points in Figure 45, which represent performance during presentations of a 30% sample. As one would expect, subjects' responses remain relatively stable under these conditions, regardless of whether a non-roving or roving stimulus paradigm was adopted. When training on the roving task began, additional conditions were presented, in which the sample contrast varied but the test contrast remained the same (grey-coloured markers). Over the course of training on the roving task, subjects' responses diverged depending on the contrast of the sample.

The slope of the best-fit line to the data was calculated for each sample contrast condition, to provide a measure of the amount of change observed over the course of training on the roving task (Table 18). One would expect that if the subjects failed to attend to the sample contrast, then the slopes would be similar across sample contrasts. On the other hand, if they modified their behaviour over the course of training, and heeded the sample contrast, then the proportion of trials in which they reported a higher test contrast (and thus the slopes of the best-fit line) would differ, depending on the sample contrast.

In 9/12 cases, the absolute value of the slope was higher for the 20% or 40% sample condition, than for the 30% sample condition. This indicated that the subjects tended to adjust their behaviour when the task called for it (during the response conflict conditions, in which the sample contrast was 20% or 40%), compared to when they had little reason to do so (during familiar conditions in which the sample contrast was 30%).

		Slope					
Test contrast (%)		22	25	28	32	35	38
Sample contrast (%)		Monkey 1					
	20	0.08	-0.25	-0.204	-	-	-
	30	0.0148	-0.234	-0.436	-0.179	-0.248	0.00881
	40	-	-	-	-0.521	-0.414	-0.398
		Monkey 2					
	20	0.914	1.83	1.359	-	-	-
	30	0.85	0.761	0.415	0.328	1.549	-0.266
	40	-	-	-	0.0333	-0.802	-0.393

Table 18. Slopes of the best-fit line to the roving data, shown in Figure 45, for each response conflict condition. The absolute value of the slope provided a measure of the degree to which performance changed over the course of training on the roving task.

In summary, when a roving stimulus paradigm was implemented, the monkeys learnt to adjust their behaviour, although this depended somewhat on which sample contrast had been presented. These results make it unlikely that subjects ignored the sample during training under roving conditions. Note that this portion of the analysis was not intended as a demonstration of perceptual learning of contrast discrimination per se, but rather, as evidence that the animals learnt to carry out their comparisons between stimuli correctly.

2.4.3 Perceptual learning averaged across test contrast conditions

Rates of learning were examined across all 12 test contrast conditions per sample contrast, using three indicators of performance for each session: 1) the mean proportion of correct responses, 2) the slope, and 3) the PSE of the psychometric curve (Figure 46).

Task performance was compared between the first and last 30% of sessions for each sample contrast. In monkey 1, when the sample stimulus had a contrast of 40%, performance improved significantly across all three measures (Table 19, unpaired two-sample *t*-test). An increase in the PSE away from 30% (corresponding to a worsening in performance) occurred for the 30% contrast sample.

In monkey 2, when the sample stimulus had a contrast of 20%, significant improvement was seen in the proportion of correct responses, the slope, and the PSE. A

trend towards a shift in the PSE towards 40% occurred, for the conditions where the sample contrast was 40%, while a significant decrease (corresponding to a worsening) in the PSE occurred for the 30% sample contrast.

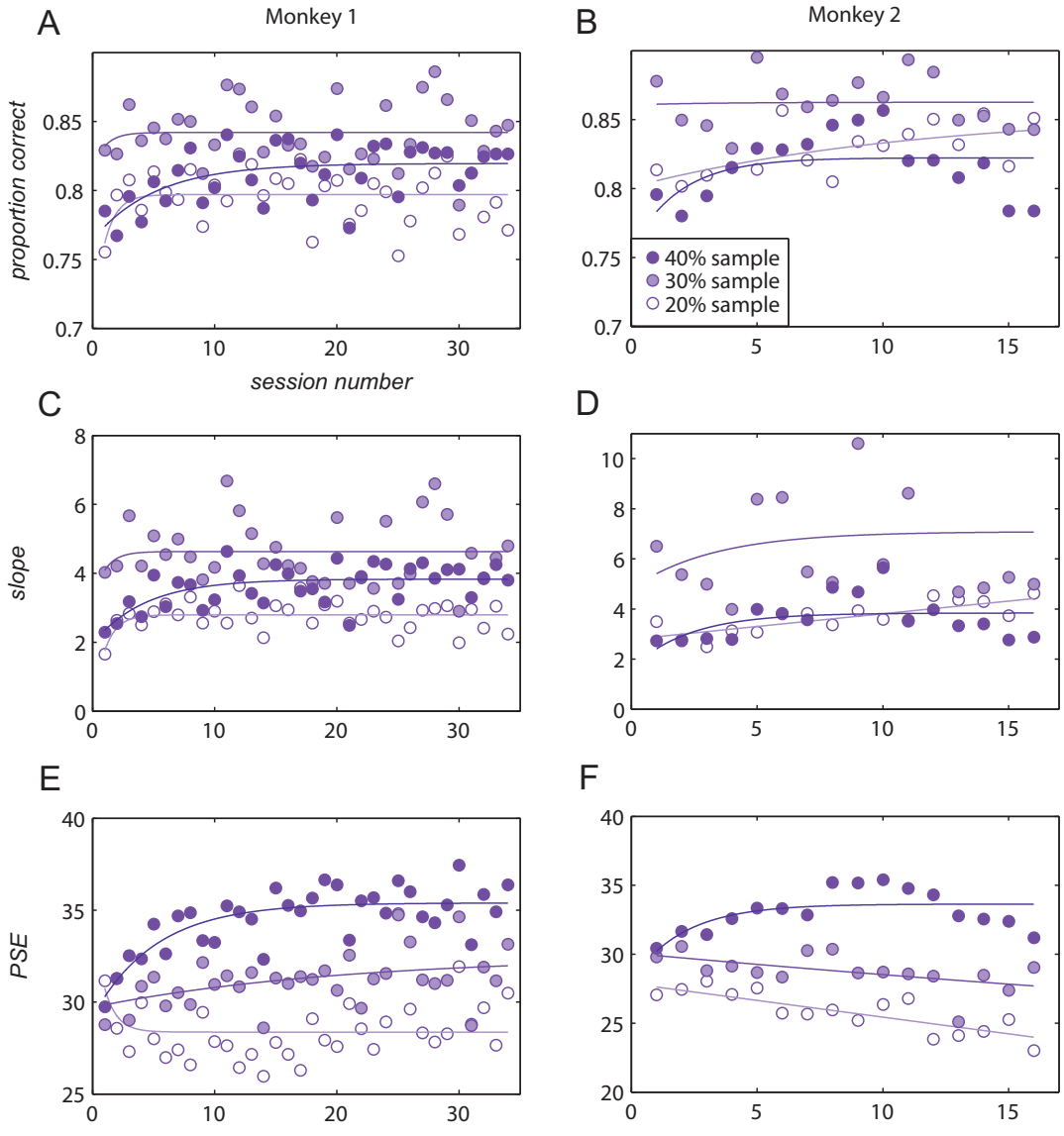


Figure 46. Performance indicators on the contrast discrimination task, over the course of roving task training (pre-flankers). A & B: $P_{correct}$; C & D: slope of the psychometric function; E & F: PSE of the psychometric function. Unfilled markers: 20% sample contrast conditions; medium-coloured filled markers: 30%; dark-coloured filled markers: 40%.

Monkey										
	1					2				
Statistic	μ_{early}	σ^2_{early}	μ_{late}	σ^2_{late}	q	μ_{early}	σ^2_{early}	μ_{late}	σ^2_{late}	q
	20% sample									
$P_{correct}$	79.4	3.5	79.2	5.0	.809	81.0	0.4	83.8	3.1	.0230*
Slope	2.7	0.2	2.7	0.2	.741	3.0	0.2	4.2	0.1	.00407*
PSE	28.3	2.1	29.6	4.9	.150	27.4	0.2	24.4	0.4	< .001*
$RT_{correct}$	149.3	125.5	112.6	53.1	< .001*	165.9	10.3	162.5	18.0	.248
RT_{error}	154.7	94.1	127.4	128.6	< .001*	169.5	16.1	169.8	14.0	.923
	30% sample									
$P_{correct}$	83.8	2.1	84.4	8.8	.563	85.1	4.1	85.7	3.4	.633
Slope	4.5	0.3	4.7	1.4	.612	5.2	1.1	7.2	21.2	.426
PSE	30.4	1.2	31.9	3.3	.0403*	29.6	0.6	27.3	2.5	.0443*
$RT_{correct}$	149.8	114.7	111.5	40.3	< .001*	165.0	7.0	162.7	23.9	.447
RT_{error}	152.9	98.1	126.2	96.7	< .001*	171.1	57.2	166.3	10.7	.288
Location	40% sample									
$P_{correct}$	79.6	3.4	82.1	1.6	.0026*	79.6	2.1	80.8	2.9	.344
Slope	3.1	0.3	3.9	0.2	.0011*	2.8	0.0	3.4	0.2	.0525
PSE	32.9	2.5	35.3	1.5	.00125*	31.5	0.8	33.0	0.8	.0542
$RT_{correct}$	148.8	123.8	113.2	52.3	< .001*	165.4	42.4	163.2	33.0	.631
RT_{error}	149.5	123.3	111.2	180.4	< .001*	168.2	6.3	160.8	30.9	.0501

* $q < \alpha$

Table 19. Comparisons of performance levels between early and late sessions during training with roving stimuli, using an unpaired *t*-test (FDR correction for α-levels, proportion correct: α =.05 × 4/4 = .05; slope: α =.05 × 4/4 = .05; PSE: α =.05 × 1/4 = .0125; $RT_{correct}$: α =.05 × 3/4 = .0375; RT_{error}: α = .05 × 4/4 = .05).

The previous figure portrayed changes in performance over the full range of test contrast conditions, regardless of task difficulty. Since subjects had already undergone extensive training during the non-roving task, we hypothesised that learning would be most evident for the response conflict conditions, whereas it would have already have reached asymptotic levels for the easy conditions. Hence, values of $P_{correct}$, as shown in Figure 46, could not convey the amount of learning that occurred for the hardest test contrast conditions.

To obtain an overview of the degree of improvement attained for these test contrasts, $P_{correct}$ was calculated based on subjects' performance during conflict conditions only. This involved taking the mean of the proportion of correct trials across all the conflict conditions for each day, regardless of sample contrast. $P_{correct}$ for the response conflict conditions was then plotted as a function of time (Figure 47). This yielded a clearer picture of the sizeable amount of learning that occurred under these conditions, for both monkeys.

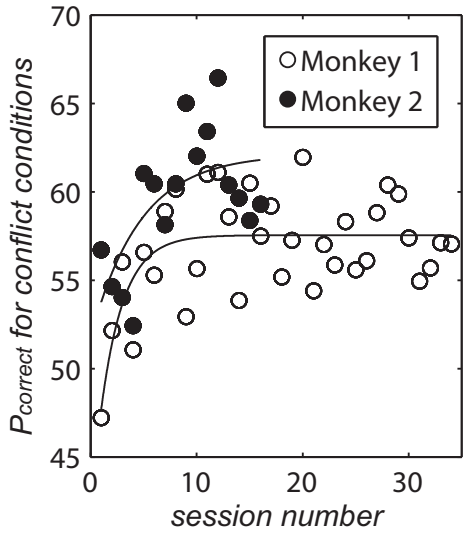


Figure 47. $P_{correct}$ (calculated based on the proportion of correct trials during response conflict conditions only), as a function of time.

2.4.4 Relative changes in performance based on sample contrast

To investigate whether improvements associated with a particular sample contrast came at the expense of performance with a different sample contrast, the proportion of correct trials was plotted between each pair of sample contrast conditions (20% versus 30% samples, 30% versus 40% samples, and 20% versus 40%, see Figure 48). If improvements in performance for a particular sample contrast were accompanied by a worsening in performance for other sample contrasts, then one would expect to observe a negative relationship between performance levels for each pair of conditions that included the sample contrast for which improvements were seen (i.e. for the 20-versus-40 and the 30-versus-40% comparisons in monkey 1, and for the 20-versus-30 and the 20-versus-40% comparisons in monkey 2).

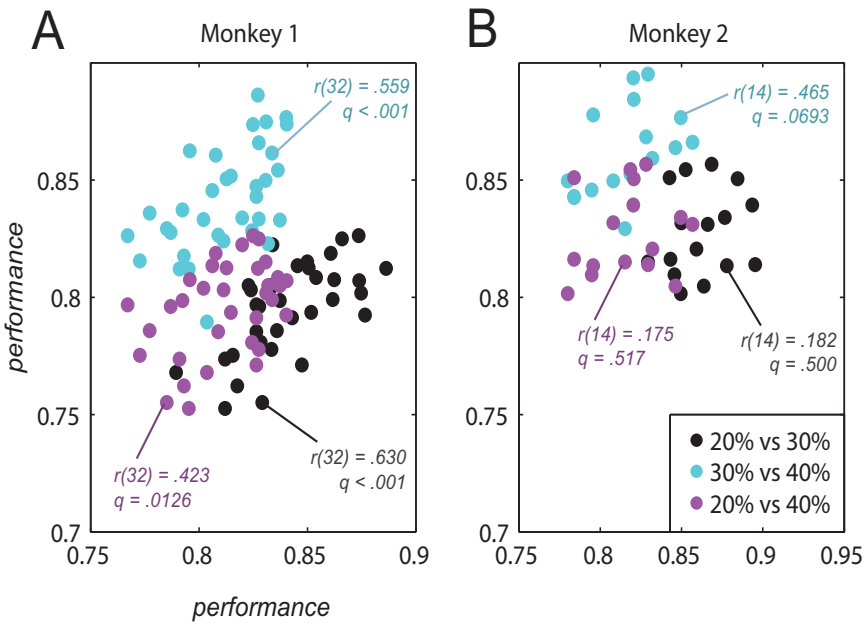


Figure 48. Proportion of correct trials, for each pairwise comparison of sample contrast conditions, for monkey 1 (A) and monkey 2 (B). 20% versus 30%: black; 30% versus 40%: cyan; 20% versus 40%: magenta.

Contrary to the above prediction, the proportions of correct trials were significantly positively correlated (Spearman's rank correlation) for each of the three comparisons made in monkey 1, while no significant relationship was observed in monkey 2 (FDR correction for multiple comparisons, $\alpha = .05/6\times3 = .025$). Thus,

improvements for selected sample contrasts did not occur at the expense of performance on other sample contrasts.

2.4.5 Psychometric thresholds during the roving task

A Spearman's rank correlation analysis was carried out between threshold and session number, to test for changes in threshold over time. Significant decreases in upper and lower threshold values were observed in monkey 1 for the 40% sample contrast and in lower thresholds in monkey 2 for the 20% and 30% sample contrasts (Table 20). These changes matched the improvements seen in subjects' performance and in the slope and PSE of their psychometric functions with training.

Statistic	*df*	*r*	*q*	*df*	*r*	*q*
	Monkey 1			Monkey 2		
			20%			
C_L	32	-.107	.545	14	.765	< .001*
C_H	32	.371	.0315	14	-.541	.0327
			30%			
C_L	32	-.303	.0814	14	.359	< .001*
C_H	32	.246	.160	14	-.406	.0327
			40%			
C_L	32	-.429	.0120*	14	-.018	.952
C_H	32	.428	.0115*	14	.356	.176

* $q < \alpha$

Table 20. Changes in psychometric thresholds during the roving task. FDR correction for multiple comparisons, $\alpha = .05/12 \times 4 = .0167$.

2.4.6 Reaction times

For each session, mean RTs were calculated separately for correct and incorrect trials, across all 12 test contrast conditions per sample contrast. We investigated whether the mean RT changed over the course of training. In monkey 1, RTs decreased significantly with training across all sample contrast conditions, for correct as well as for incorrect trials, while in monkey 2, a significant reduction in RT occurred only during training with the 40% sample contrast, for incorrect trials (Table 21).

	Monkey 1		Monkey 2	
Statistic	*r*	*q*	*r*	*q*
	20% contrast			
Correct	-.9089	< .001*	-.3798	.1468
Error	-.7678	< .001*	-.0818	.7633
	30% contrast			
Correct	-.9127	< .001*	-.3432	.1931
Error	-.8126	< .001*	-.3622	.1681
	40% contrast			
Correct	-.9051	< .001*	-.2627	.3257
Error	-.8709	< .001*	-.5914	.0158*

* $q < \alpha$

Table 21. Pearson's correlation coefficients and *q*-values for correlations between mean RT and session number. FDR correction, $\alpha = .05/12 \times 7 = .0292$.

2.4.7 Discussion of behavioural changes during the roving task

Our data show that under roving conditions, perceptual learning occurred in both monkeys, although the changes differed slightly between the two animals (e.g. improvements in the PSE occurred for different sample contrasts between the monkeys). We found that performance with a sample of 30% contrast remained comparable to that observed prior to training on the roving task, indicating that the improvements observed for the 40% and 20% sample conditions in monkeys 1 and 2, respectively, did not occur at the expense of previous improvement.

Adini et al. (2004) did not observe improvements in performance when naïve observers were trained on the MBT roving task. Neither did they see improvements among subjects who had previously received training on a blocked task before embarking on the MBT task. In Yu et al.'s MBT roving task (2004), naïve subjects delivered results that varied across individuals (as described earlier in this chapter). The regimen followed by our subjects differed slightly from each of these groups of human subjects, as our monkeys were exposed to a non-roving task, followed by what was presumably the maximally challenging version of the roving task (the MBT method). Moreover, to keep the task manageable for our monkeys, we used three sample contrasts, whereas Yu et al. (2004) used four reference contrasts, and Adini et al. (2004) used seven.

Nonetheless, on the whole, our observations matched those seen in previous studies with human subjects (Yu et al. (2004) and Adini et al. (2004)): improvement was possible albeit to a limited degree; it took place only under a subset of conditions; and inconsistencies occurred between subjects.

2.5 Roving task neuronal results

2.5.1 Changes in the CRF during training on the roving task

2.5.1.1 Individual channel results

Significant changes in several of the four measures of the contrast response function were observed in a number of channels. The slope became shallower for 9/25 channels in monkey 2 (the breakdown by sample contrast is shown in Table 22), and the C_{50} shifted away from C_{sample} when the sample contrast was 30%, in 10/25 channels in monkey 2. These changes did not appear to be closely linked to the improvements observed at the behavioural level (better performance with the 40% sample contrast in monkey 1, and with the 20% sample contrast in monkey 2). Thus, behavioural improvements could not be adequately explained by a change in the CRFs of individual V1 channels during the roving task.

		Monkey 1			Monkey 2		
Sample contrast (%)		20	30	40	20	30	40
Slope versus session	Steeper	0	0	0	2	0	2
	Shallower	9	11	11	3	3	3
C_{50} versus session	Towards C_{sample}	1	2	0	1	0	1
	Away from C_{sample}	1	2	2	1	7	2
Minimum versus session	Increase	0	0	0	1	0	1
	Decrease	0	0	0	0	0	0
Maximum versus session	Increase	1	0	0	2	4	1
	Decrease	3	3	2	0	3	2

Table 22. Number of channels with significant changes in each parameter of the contrast response function, during training with roving sample stimuli (monkey 1: $N = 23$; monkey 2: $N = 25$).

2.5.1.2 Population results

To identify changes in the CRF during training with roving sample stimuli, population CRFs were calculated in the same way as that reported in Chapter 1 during the non-roving task. The four parameters obtained from the fitted curves were plotted against session number (Figure 49).

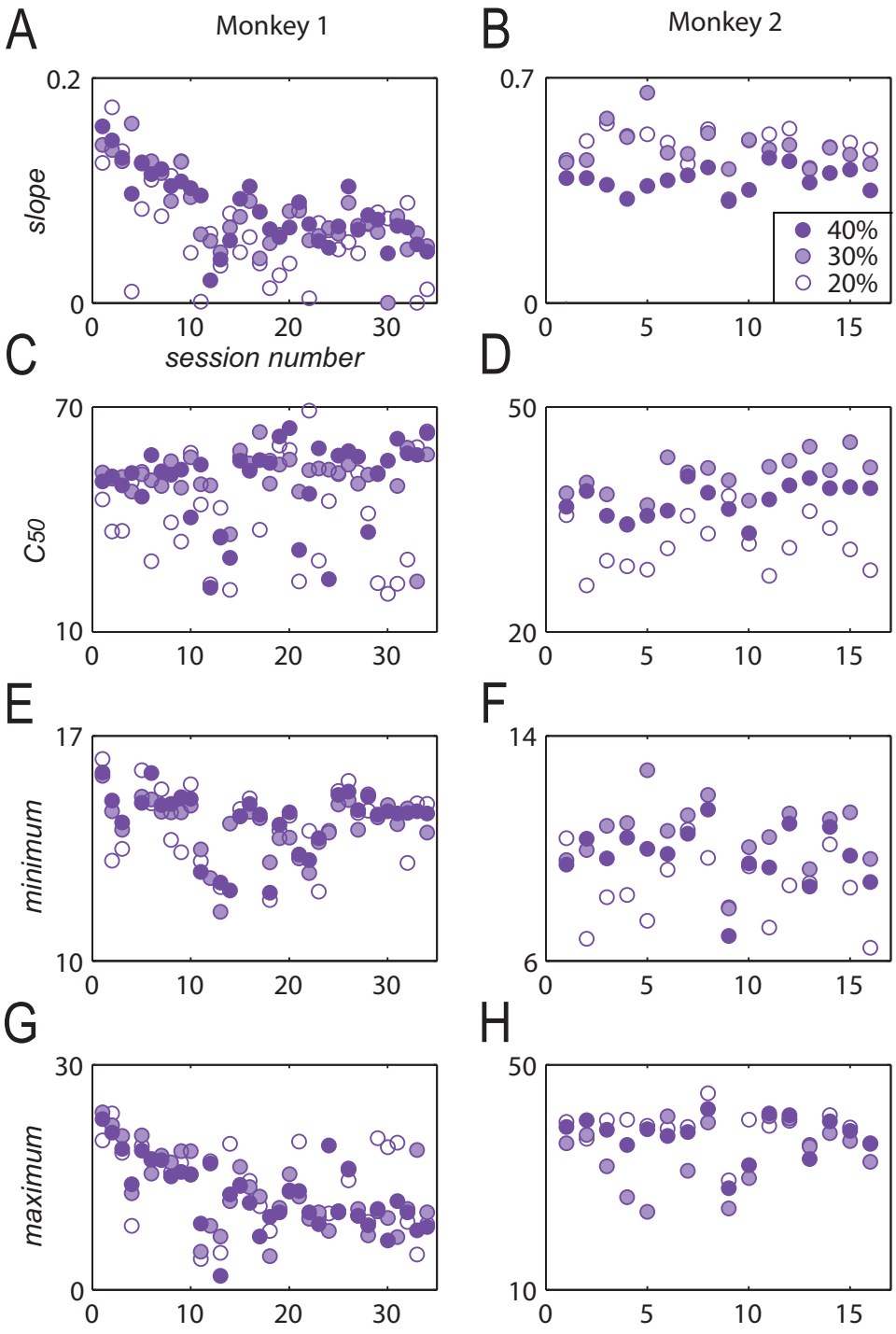


Figure 49. Parameter values of the population CRF with time, during training with roving sample stimuli. Left column: monkey 1; right column: monkey 2. A & B: slope; C & D: C_{50}; E & F: minimum value; G & H: maximum value. Unfilled markers: 20% sample; medium: 30%; dark: 40%.

During training with roving stimuli, significant decreases in the slope were observed amongst population responses in monkey 1 for the 30% and 40% sample conditions, and a non-significant decrease was seen for the 20% sample (Spearman's rank correlation, Table 23). Simultaneously, a decrease in the maxima was observed in monkey 1 for the 30% and 40% sample contrasts. No changes were seen in either the slope or the maxima for monkey 2. These observations did not appear to correlate closely with the changes seen at the behavioural level- monkey 1's performance improved for the 40% sample condition, while monkey 2's performance improved for the 20% sample condition. If neuronal changes had matched psychometric changes, one might expect to see a steepening of the CRF for the 40% and 20% sample contrast for the respective monkeys. Instead, roving training was marked by either no change or a decrease in the slope of the CRF.

	Spearman's rank correlation								
Sample contrast (%)	20			30			40		
Statistic	*df*	*r*	*q*	*df*	*r*	*q*	*df*	*r*	*q*
	No flankers								
	Monkey 1								
Slope	32	-.351	.0425	32	-.646	< .001*	32	-.668	< .001*
C_{50}	32	.058	.743	32	.234	.182	32	.446	.00822*
Min	32	-.246	.160	32	-.278	.112	32	-.218	.215
Max	32	-.336	.0526	32	-.540	.00117*	32	-.642	< .001*
	Monkey 2								
Slope	14	-.235	.379	14	-.282	.288	14	.200	.456
C_{50}	14	.103	.705	14	.662	.00654*	14	.518	.0423
Min	14	-.062	.822	14	.035	.900	14	-.159	.556
Max	14	-.144	.594	14	.221	.410	14	-.129	.633

* $q < \alpha$

Table 23. Descriptive statistics for a Spearman's rank correlation analysis to identify changes in the slope, C_{50}, and minimum and maximum values of the CRF, during training with roving stimuli. Significant decreases in slope and the maxima occurred for monkey 1, for the 30% and 40% sample contrast conditions (FDR correction, $\alpha = .05/24\times6 = .0125$).

The C_{50} increased significantly away from 40% for the 40% sample in monkey 1; it also increased significantly away from 30% for the 30% sample in monkey 2. Although the shift seen in monkey 1 resulted in the movement of the C_{50} away from the

sample contrast, this may still have served to separate the neuronal responses to the 40% sample from those to the 20% and 30% sample. In monkey 2, however, this did not appear to be the case, as the C_{50} for the 30% sample became higher than the C_{50}s for both the 20% and the 40% samples.

2.5.2 Changes in PROBMAT during training with roving stimuli

2.5.2.1 Individual channel results

Neuronal data from individual channels were monitored for changes in the PROBMAT function over the course of training with roving samples. Most of the changes consisted of decreases in slope in monkey 1, for the 20% and 30% sample contrasts (Table 24). This apparent drop in the discriminability of neuronal responses was reminiscent of the decreases observed in maximum firing rates, as reported in the section on changes in the CRF of individual channels (page 136). However, these changes were unable to account for the improvements seen at the behavioural level (with the 40% sample contrast conditions in monkey 1 and the 20% sample contrast in monkey 2).

		Monkey 1			Monkey 2		
		20	30	40	20	30	40
Slope versus session	Steeper	0	0	0	0	0	1
	Shallower	8	7	2	2	1	0
PNE versus session	Towards C_{sample}	0	0	0	2	0	1
	Away from C_{sample}	0	0	0	1	0	0
Minimum versus session	Increase	0	0	0	0	0	0
	Decrease	0	0	0	0	0	0
Maximum versus session	Increase	1	1	0	2	0	1
	Decrease	0	1	0	0	0	0

Table 24. Number of channels with significant changes in each parameter of the PROBMAT-versus-contrast function, during training with roving sample stimuli (monkey 1: $N = 23$; monkey 2: $N = 25$, FDR correction for multiple comparisons).

2.5.2.2 Population results

To identify changes in the PROBMAT values during the training period, the population PROBMAT-versus-contrast function was plotted for each session, and the

four parameters obtained from the fitted curves were plotted against session number (Figure 50).

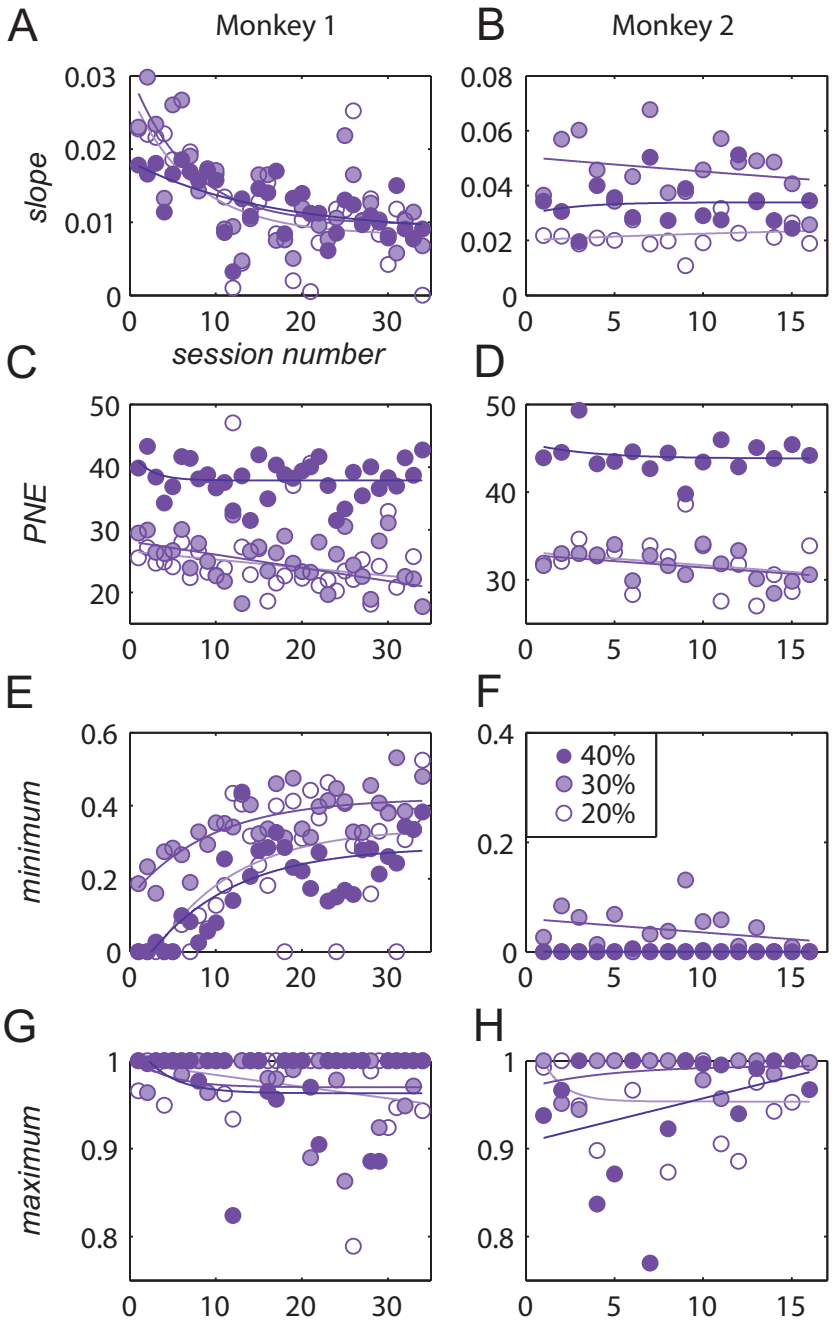


Figure 50. Parameter values of the population PROBMAT curve during training with roving stimuli at the V1 location. Left column: monkey 1; right column: monkey 2. A & B: slope; C & D: PNE; E & F: minimum value; G & H: maximum value. Unfilled markers: 20% sample; medium purple: 30%; dark purple: 40%. Significant decreases in the slope and increases in the minimum values were seen for all three sample contrasts for monkey 1, while no changes were observed in monkey 2 (see Table 25).

Significant decreases in the slope and increases in the minimum were observed for the population responses in monkey 1 during training with roving stimuli, for all three sample contrast conditions (Table 25). Unlike the changes seen in performance at the behavioural level, no particular improvement was observed in the PNE for the 40% sample.

Sample contrast (%)	Spearman's rank correlation								
	20			30			40		
Statistic	*df*	*r*	*q*	*df*	*r*	*q*	*df*	*r*	*q*
	Monkey 1								
Slope	32	-.609	< .001*	32	-.554	< .001*	32	-.640	< .001*
PNE	32	-.361	.0366	32	-.398	.0205	32	-.006	.972
Min	32	.545	.00102*	32	.669	< .001*	32	.698	< .001*
Max	32	.099	.574	32	.065	.716	32	-.010	.957
	Monkey 2								
Slope	14	.182	.498	14	-.103	.705	14	-.071	.797
PNE	14	-.265	.321	14	-.468	.0698	14	.044	.874
Min	14	.080	.770	14	-.447	.0844	14	.112	.681
Max	14	-.138	.609	14	.015	.961	14	.397	.129

* $q < \alpha$

Table 25. Statistics for a Spearman's rank correlation analysis to identify changes in the slope, PNE, and minimum and maximum values of the neurometric function, during training on roving stimuli. Significant decreases in slope and increases in the minimum value were seen in monkey 1 for all three sample contrast conditions (FDR correction, $\alpha = .05/24\times6 = .0125$).

2.5.3 Neurometric thresholds during the roving task

An analysis of neurometric thresholds was carried out for data from roving sessions. Thresholds were monitored over time for training-induced changes (refer to Figure 51 and Table 26). Significant shifts were observed in 4/6 cases in monkey 1; however, in none of these cases did the change correspond to an improvement in threshold (Spearman's rank correlation). No changes in threshold were observed in monkey 2.

These observations, at the neuronal level, did not match the pattern of improvement seen at the behavioural level, in which significant improvements in psychometric thresholds occurred for upper thresholds in monkey 1 for the 40% sample

contrast and in lower thresholds in monkey 2 for the 20% and 30% sample contrasts (page 133).

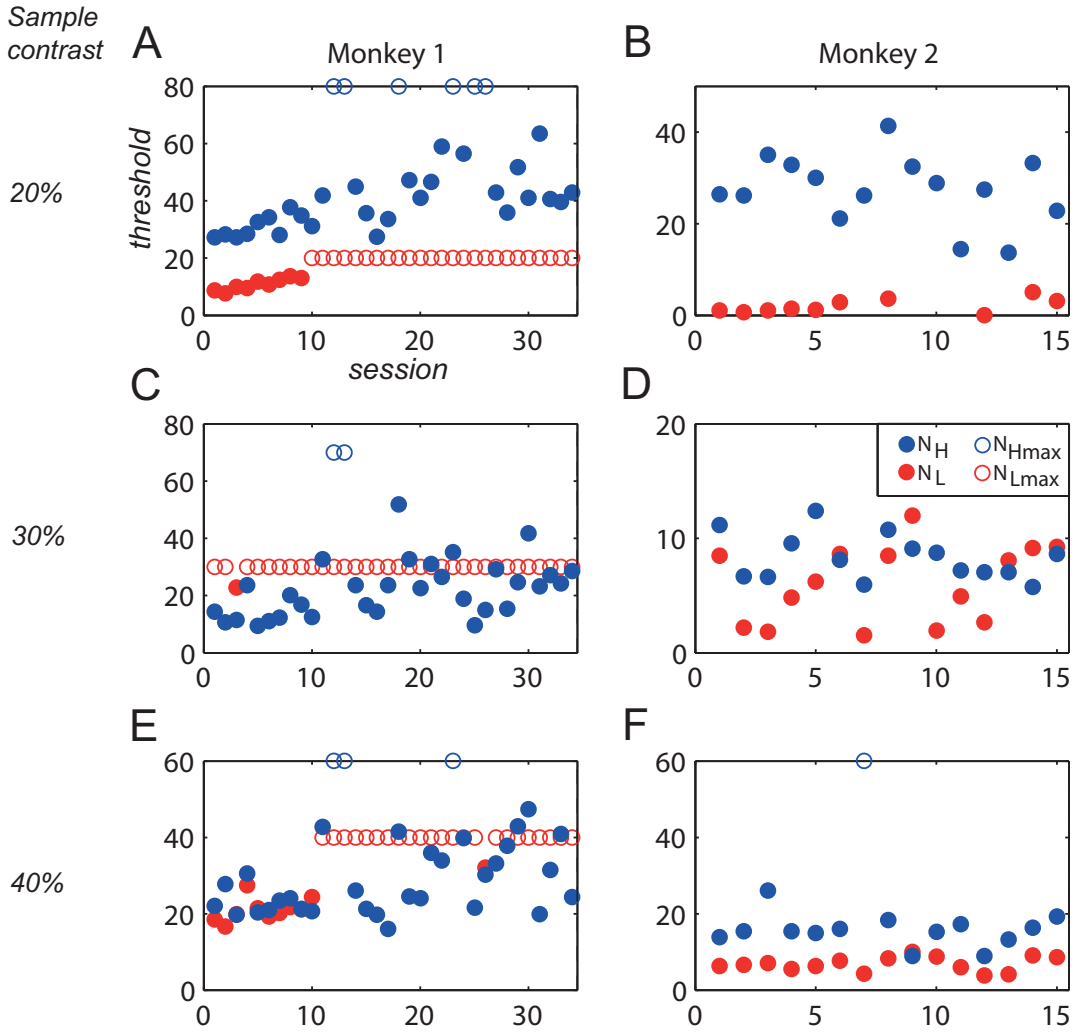


Figure 51. Neurometric thresholds (filled markers), plotted as a function of time, during training on a roving stimulus task. Unfilled markers indicate sessions where thresholds could not be obtained. Left column: monkey 1; right column: monkey 2. Top row: 20% sample; middle row: 30% sample; bottom row: 40% sample. Red markers: N_L conditions (the test contrast was lower than that of the sample); blue markers: N_H conditions (the test contrast was higher than that of the sample). In a number of cases, thresholds grew significantly worse for monkey 1 (refer to Table 26 for results from the correlation analysis).

Statistic	*df*	*r*	*q*	*df*	*r*	*q*
	Monkey 1			Monkey 2		
			20%			
C_L	32	.775	< .001*	14	.124	.648
C_H	32	.573	< .001*	14	-.079	.771
			30%			
C_L	32	.257	.142	14	.478	.648
C_H	32	.425	.0122*	14	-.052	.771
			40%			
C_L	32	.753	< .001*	14	.032	.908
C_H	32	.367	.0329	14	-.150	.579

* $q < \alpha$

Table 26. Spearman's rank correlation coefficients and *q*-values, indicating changes in threshold over the course of training with roving stimuli. FDR correction for multiple comparisons for flankerless data: $\alpha = .05/12 \times 4 = .0167$.

2.5.4 Variability of the visual response during the roving task

To examine changes in variability of the spike response during training with roving stimuli, the *FF* was monitored across sessions, in the same manner as that reported in the methods section of Chapter 1 (page 50). A two-factor ANOVA was performed to identify a main effect of training period, for each channel.

Significant changes in the *FF* were seen on a number of channels for monkey 1 (20% sample: 16/23 channels, 30%: 12/23 channels, 40% sample: 16/23 channels), and in all cases, they consisted of decreases in *FF* with training. In monkey 2, changes were observed for a small number of channels (20% sample: 6/25 channels, 0/6 decreases, 6/6 increases; 30%: 3/25 channels, 2/3 decreases, 1/3 increase; 40% sample: 7/25 channels, 6/7 decreases, 1/1 increase).

At the population level, decreases in the *FF* were observed for all three sample contrast conditions in monkey 1, and for the 40% sample in monkey 2, while increases were seen for the 20% sample in monkey 2 (two-factor ANOVA, Table 27). These results matched those seen at the individual channel level, where the *FF* decreased across multiple channels across sample conditions in monkey 1, but increased for the 20% sample condition in monkey 2.

	Subject 1			Subject 2		
Sample contrast (%)	*df*	*F*	*q*	*df*	*F*	*q*
20	1,5496	34.0	< .001*	1,2376	7.8	.00517*
30	1,5496	33.0	< .001*	1,2376	0.1	.707
40	1,5496	31.6	< .001*	1,2376	17.7	< .001*

* $q < \alpha$

Table 27. Results from two-factor ANOVA, comparing trial-wise spike variability between early and late roving sessions. Significant changes in the Fano factor occurred over the course of training, in 5/6 cases (FDR correction for multiple comparisons, $\alpha = .05/6 \times 5 = .0417$).

2.5.5 Discussion of neuronal results from the roving task

In this chapter, unlike that seen in the previous chapter on non-roving stimuli, results differed between subjects, and changes in neuronal activity, where present, were little correlated with behavioural changes (if at all).

Improvement in psychometric performance was observed in monkey 1 for the 40% sample, and in monkey 2 for the 20% sample. If underlying neuronal changes had matched those seen at the behavioural level, one would expect to observe a shift in the C_{50} and the PNE towards 40% and 20% in monkeys 1 and 2, respectively, as was seen in V4 during the non-roving task. Alternatively, if results matched those seen in V1 during the non-roving task, then overall activity levels might decline, and/or the C_{50} might shift away from 30%, as seen in monkey 2 during the non-roving task.

What we in fact observed was that activity levels decreased with training in monkey 1, across all three sample contrasts, as shown by the decrease in maxima of the CRF. This result was similar to the population decreases in V1 activity that were observed by Ghose et al. (2002), during training on an orientation discrimination task. Note also that we had previously noticed significant decreases in individual channel V1 activity in monkey 2, although this was during training on the non-roving rather than the roving task.

Furthermore, we observed an increase in the minima of the PROBMAT neurometric function in monkey 1, across all three sample contrasts, which

corresponded to a narrowing in the range and a worsening of discriminability in spiking responses to sample and test stimuli.

Spike variability decreased with training for monkey 1, for all sample contrasts, while an increase was seen for the 20% sample in monkey 2. CD task performance increased for monkey 2 on the 20% sample; however, no such increase was seen in monkey 1, despite improvements for the 40% sample, thus there did not appear to be a clear relationship between the *FF* and task performance at the behavioural level.

In summary, we failed to find evidence that V1 was principally responsible for the limited improvements seen during roving training; instead, responses in area V1 generally seemed to decrease with training. This decrease may not necessarily be a direct result of roving training, but may be a non-specific phenomenon that accompanies perceptual learning. Imaging studies in human subjects have explored the relationships between changes in blood-oxygen-level-dependent (BOLD) signal and the degree of perceptual learning achieved. Mukai et al. (2007) reported that amongst their subjects, those who improved significantly on a grating waveform discrimination task (Fiorentini, 1980) showed a gradual decrease in BOLD activity in visual and attention-related areas (18, 19, FEF, SEF, and IPS) over the course of training, whereas activity levels in these areas remained high for 'non-learners.' This result appeared to match our observations (though also note that Schwartz et al. (2002) reported an increase in activity in intensively-trained retinotopic regions within V1, compared to untrained regions, during a visual texture discrimination task).

The pattern of decreasing V1 activity may reflect changes that accompany over-training; according to the predictions made by the RHT, changes may occur in certain cortical regions during the early phase of training, and then shift to other sites as learning becomes more finely tuned. We observed behaviourally-coupled changes in V4 during the non-roving task, and subsequently observed relatively few behaviourally-dependent changes in V1 during both the non-roving and roving tasks. As training at the two locations was not carried out simultaneously, it is not possible for us to conclusively identify the exact time course of changes in neuronal activity. However, based on the changes observed at an intermediate level of the visual hierarchy, our data appear to support either the late learning model or the RHT model of PL.

Chapter 3: Flanker task

3.1 Flanker task literature review

In tasks involving flanker stimuli, a central stimulus is positioned at a specific location in the visual field; this stimulus of interest is flanked by one or more additional stimuli. In electrophysiological studies, flankers are usually used to investigate how the stimulation of regions outside the classical receptive field affects neuronal responses. In psychophysics studies, the concept of the 'receptive field' remains vaguely defined; however, it provides a useful way of framing and investigating excitatory, inhibitory and/or masking effects on stimulus processing and perception.

Human psychophysics experiments have documented a variety of results, particularly for parameters such as the distance between flankers and central target stimuli (Adini & Sagi, 2001; Polat, 1999; Polat et al., 2004; Polat & Sagi, 1993; Zenger & Sagi, 1996); the orientation (Dorais & Sagi, 1997; Yu et al., 2004), spatial frequency (Polat et al., 2004; Yu et al., 2004) and contrast of the target (Adini & Sagi, 2001; Adini et al., 2002; Adini et al., 2004; Polat, Mizobe, Pettet, Kasamatsu, & Norcia, 1998; Tsodyks et al., 2004; Yu et al., 2004); the contrast (Yu et al., 2004; Yu, Klein, & Levi, 2002; Zenger & Sagi, 1996), position (Adini & Sagi, 2001; Tsodyks et al., 2004; Yu et al., 2002), size (Saarela, Sayim, Westheimer, & Herzog, 2009), and orientation (Polat et al., 1998; Polat & Sagi, 1993; Yu et al., 2002; Zenger & Sagi, 1996) of flankers; and the length of chains of flanker stimuli (Adini & Sagi, 2001; Adini et al., 2002; Tsodyks et al., 2004).

The potential for flanker-induced improvements in performance of visual tasks has generated interest in whether the use of flanker stimuli might aid perceptual learning. A human psychophysics study by Adini et al. (2002) examined the effects of flanker training on CD thresholds, and found that while training with flankerless stimuli produced no significant improvement, the presence of flanker stimuli during training yielded reductions in threshold of ~50%.

This report elicited a series of follow-up experiments, particularly from the Yu group (Yu et al., 2004), which claimed that on the contrary, it was possible to boost performance simply by continuing the training regime for a longer period, and that the addition of flankers was unnecessary. Yu et al.'s study consisted of several components. First, in an effort to replicate Adini et al.'s findings, subjects performed a CD task with a very similar task paradigm to that used by Adini et al., where reference contrasts were presented in blocks, and ranged from 0 to 63%. Contrary to Adini et al.'s results, significant improvement was seen as a result of training, and the degree of improvement in Yu et al.'s flankerless training matched that obtained by Adini et al. after flanker training. This procedure was repeated in a second group of subjects, but subjects were allowed to continue their CD training over a larger number of sessions, until their performance reached asymptote levels. This was followed by further training in the presence of flankers, which essentially yielded no further improvement, indicating that the addition of flankers had been ineffective. Up to this point, their results closely matched those reported in Chapter 1 of this thesis, on training with a non-roving paradigm.

Their study also implemented a roving version of the CD task. Yu et al. (2004) trained subjects on a roving task, presenting flankerless Gabor stimuli under four contrast conditions at the fovea, and found no systematic improvements across subjects. The same group of subjects continued their training, this time with the addition of three pairs of flankers; no further improvement was observed. Thus, the overall conclusion was that flankers were not able to lower CD thresholds under either non-roving or roving conditions.

Tsodyks et al. (2004) compared performance after training in either the absence or presence of flanker Gabors. The researchers observed a contrast-dependent 'masking effect'- for very low-contrast targets, flankers improved thresholds (as in Polat and Sagi (1993)), but this effect was reversed for target contrasts that were higher than the detection threshold. Next, they found that while the length of flanker chains had no effect on the levels of facilitation seen with sub-detection-threshold targets, flanker chain length did have an effect for supra-detection-threshold targets.

In another study by Adini et al. (2004), subjects were initially trained on both a non-roving CD task (with 7 pedestal contrast conditions, presented in blocks) and a roving task. After flanker training, improvements were seen in the threshold, particularly when trials were blocked by pedestal contrast, and improvement took place to a small extent when flanker practise was carried out with roving stimuli. The authors suggested that the variation in flanker chain length during training may account for the learning effects observed; furthermore, they pointed out that Yu et al. (2004) had used a slightly different flanker stimulus (a fixed-length elongated Gabor) instead of chains of Gabors, thus this factor may have contributed to the lack of learning seen in Yu et al.'s study.

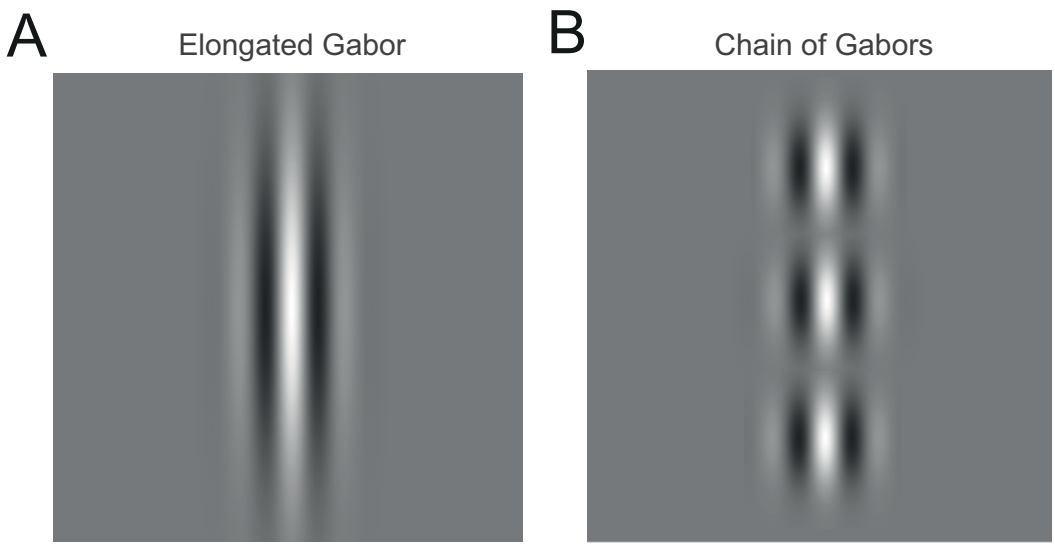


Figure 52. Illustration of the difference between (A) the elongated Gabor stimuli used by Yu et al. (2004), and (B) the chains of Gabor stimuli used by Adini et al. (2004).

In summary, a number of studies have demonstrated flanker-induced improvements in contrast discrimination under roving conditions. While this was not replicated across all the groups, we were still intrigued by the possibility of boosting performance on the roving task above the maximum levels attained thus far by our macaque subjects. Therefore, the purpose of the next stage of training was to investigate whether the presence of additional, flanker stimuli would lead to a boost in subjects' performance on the roving task.

3.1.1 Goals of the flanker task

Based on reports from the human psychophysical literature, flanker-induced improvements might potentially occur, given favourable circumstances. Our main goal was to investigate whether the addition of flankers would trigger a surge in performance beyond the plateau seen towards the end of flankerless roving training. To optimise our chances of success, we followed Adini et al.'s paradigm (2004), using chains of Gabors (rather than the elongated Gabors used by Yu et al.) and keeping the contrast of flankers constant at 30% throughout training, regardless of the sample contrast.

However, we continued to vary the sample contrast from trial to trial, even though Adini et al. reported better results for a blocked than for an MBT method, because we wanted to keep our paradigm as similar as possible to that used in the previous stage of roving training and ensure a smooth transition to the flanker task for our monkeys.

Neuronal activity was monitored throughout flanker training. As mentioned in the previous chapter, our monitor screens were not able to accommodate flankers at the V4 location due to the large size of the V4 RFs and their distance from the centre of vision, thus this stage of training was carried out solely with stimuli positioned at the V1 location.

3.1.1.1 Psychophysics/ behavioural questions

- Upon the addition of flanker stimuli, do macaque subjects show further improvements in contrast discrimination?
- If so, are these improvements seen across numerous sample contrasts, or only for select ones?

3.1.1.2 Neurophysiological questions

- Do changes in spiking activity occur in area V1?
- What is the nature of these changes (e.g. alterations of firing rate, spike variance, and tuning properties)?
- Do neurometric and behavioural changes correlate with each other?

3.2 Methods

3.2.1 Stimuli used in the flanker task

In addition to the sample and test grating stimuli, flanker gratings were displayed collinearly, immediately above and below the central stimuli, forming a column of three gratings that were positioned edge to edge (Figure 53). The flanker stimuli were identical to the sample and test stimuli in terms of size, SF and orientation, while their contrast was kept fixed at 30% throughout training.

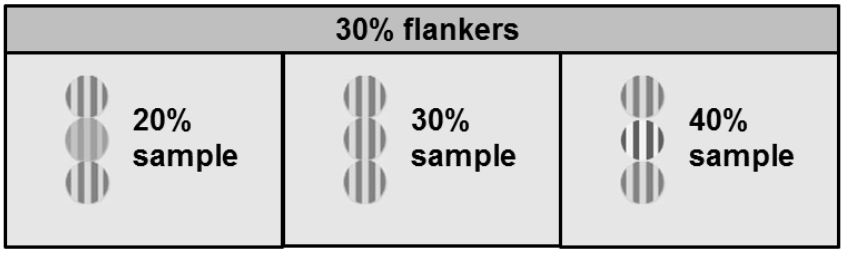


Figure 53. Relative positions of stimuli used during the flanker task (contrast levels are exaggerated for illustrative purposes).

3.2.2 Measures of perceptual learning

As with the previous analyses carried out with data from the flankerless roving task, levels of psychophysical and neurometric performance were monitored throughout training.

3.3 Flanker task behavioural results

3.3.1 Training on a roving task with flankers at the V1 location

Subjects practised a roving contrast discrimination task with flanker stimuli for several weeks, until their performance reached a plateau (monkey 1: 15 sessions, spanning a period of 6 weeks; monkey 2: 22 sessions, spanning 6 weeks). As with the flankerless paradigm in the previous section, learning rates were monitored across all 12 test contrast conditions per sample contrast, using three measures of performance for

each session (Figure 54 and Figure 55).

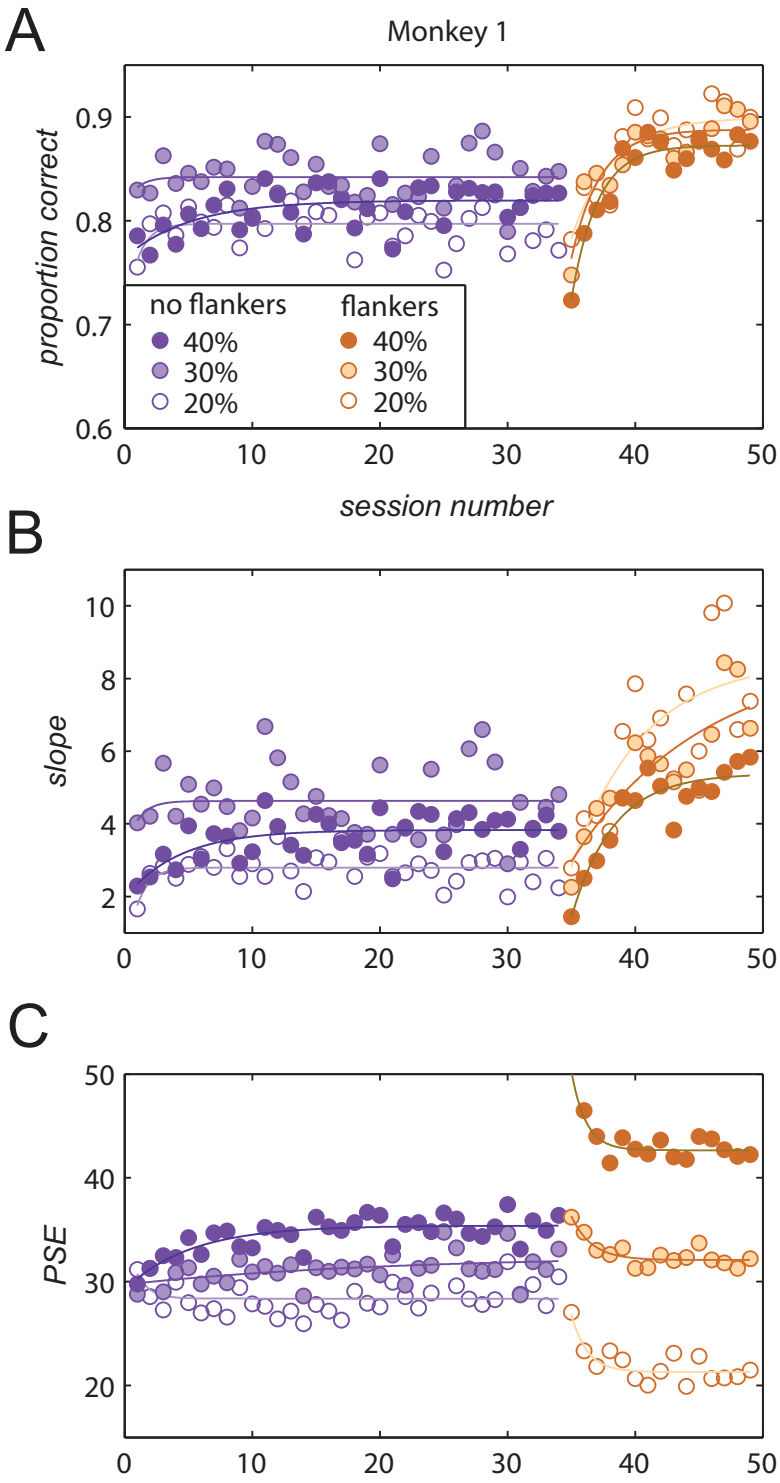


Figure 54. Performance levels in monkey 1, during training in the presence of flanker stimuli (orange), at the V1 location. Performance levels prior to the addition of flankers (purple) are replicated from Figure 46 (page 129).

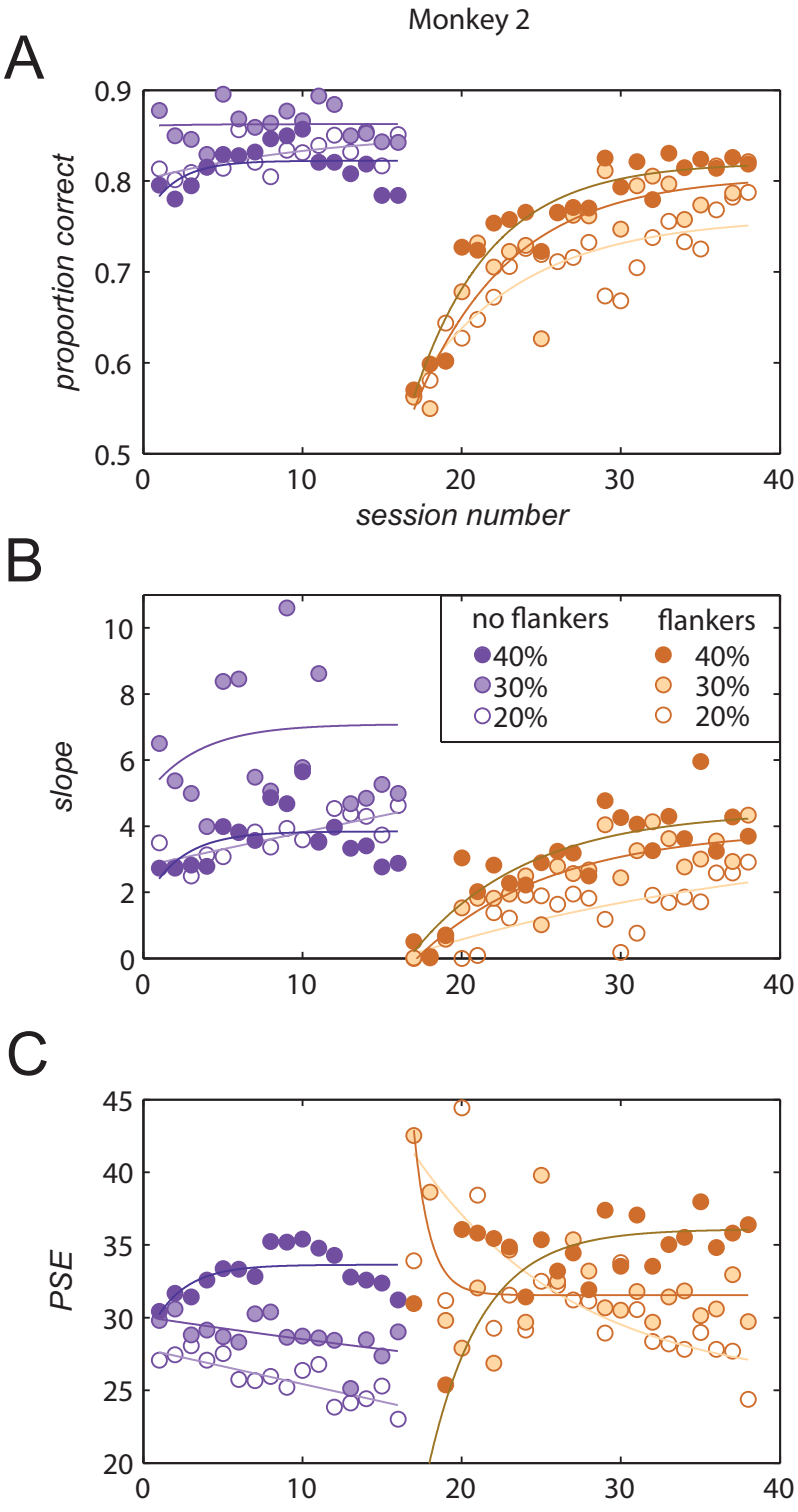


Figure 55. Performance levels in monkey 2, during training in the presence of flanker stimuli (orange), at the V1 location. Performance levels prior to the addition of flankers (purple) are replicated from Figure 46.

Task performance was compared between the first and last 30% of flanker sessions, for each sample contrast. The proportion of correct trials and the slope increased significantly for monkey 1, across all sample contrasts, while the PSE shifted significantly towards the sample contrast, when the sample contrast was 20% and 30% (refer to Table 28). For monkey 2, significant improvement was seen in the proportion of correct trials and the slope, for all three sample contrasts, and a shift of the PSE occurred towards the value of 20%, for the 20% sample contrast condition. Thus, during the period of training with flanker stimuli, improvements were seen in both subjects, across all sample contrast conditions, whereas with previous training in the absence of flankers, improvements were only seen for a limited subset of sample contrasts.

Monkey	1			2		
Statistic	*df*	*t*	*q*	*df*	*t*	*q*
	20% sample					
$P_{correct}$	1,6	30.2	.00152*	1,10	46.4	< .001*
Slope	1,6	25.6	.00231*	1,10	34.2	< .001*
PSE	1,6	6.7	.0408*	1,10	8.7	.0145*
$RT_{correct}$	1,6	0.0	.876	1,10	3.0	.113
RT_{error}	1,6	0.0	.959	1,10	13.7	.00414*
	30% sample					
$P_{correct}$	1,6	12.7	.0118*	1,10	21.7	< .001*
Slope	1,6	23.8	.00278*	1,10	31.7	< .001*
PSE	1,6	7.4	.0351*	1,10	0.5	.498
$RT_{correct}$	1,6	0.3	.584	1,10	10.0	.0102*
RT_{error}	1,6	0.0	.961	1,10	4.9	.0517
	40% sample					
$P_{correct}$	1,6	15.3	.00792*	1,10	23.0	< .001*
Slope	1,6	33.4	.00117*	1,10	16.8	.00215*
PSE	1,6	2.2	.191	1,10	1.9	.193
$RT_{correct}$	1,6	0.0	.870	1,10	1.2	.298
RT_{error}	1,6	0.1	.796	1,10	0.8	.403

* $q < \alpha$

Table 28. Comparisons of performance between early and late sessions in the presence of flankers, using an unpaired *t*-test. (Student's *t*-test, FDR correction for α-levels, proportion correct: α =.05 × 4/4 = .05; slope: α =.05 × 4/4 = .05; PSE: α =.05 × 1/4 = .0125; $RT_{correct}$: α =.05 × 3/4 = .0375; RT_{error}: α = .05 × 4/4 = .05).

An important question was whether the improvements seen within the flanker training period resulted in performance levels that surpassed those seen prior to the

addition of flankers. A comparison of levels of performance between pre-flanker and flanker training revealed that indeed, for monkey 1, the gains made during training in the presence of flankers placed his performance above that attained in the absence of flankers (Figure 54). Values of $P_{correct}$ and the slope were significantly higher by the end of flanker training, than at the end of pre-flanker training, for all three sample contrast conditions (monkey 1, 20% sample: $P_{correct}$, $t(1,12) = 65.9$, $q < .001$, slope, $t(1,12) = 108.7$, $q < .001$; 30% sample: $P_{correct}$, $t(1,12) = 13.0$, $q = .00363$, slope, $t(1,12) = 15.8$, $q = .00184$; 40% sample: $P_{correct}$, $t(1,12) = 49.1$, $q < .001$, slope, $t(1,12) = 42.4$, $q < .001$, Student's *t*-test, FDR correction for α-levels, proportion correct: $\alpha = .05 \times 4/4 = .05$; slope: $\alpha = .05 \times 4/4 = .05$). Improvements in the PSE also occurred for sample contrasts of 20% and 40% (monkey 1, 20% sample: PSE, $t(1,12) = 57.6$, $q < .001$; 30% sample: PSE, $t(1,12) = 0.01$, $q = .919$; 40% sample: PSE, $t(1,12) = 122.5$, $q < .001$, Student's *t*-test, FDR correction for α-levels, PSE: $\alpha = .05 \times 1/4 = .0125$).

The pattern observed in monkey 2's performance was markedly different (Figure 55). Values of $P_{correct}$ were significantly worse at the end of flanker training than at the end of flankerless training, for sample contrasts of 20% and 30%, while no change was observed for the 40% sample (monkey 2, 20% sample: $P_{correct}$, $t(1,8) = 29.2$, $q < .001$; 30% sample: $P_{correct}$, $t(1,8) = 20.3$, $q = .00198$; 40% sample: $P_{correct}$, $t(1,8) = 3.2$, $q = .111$). The slope of the psychometric function decreased, and the PSE shifted away from the value of 20%, for the 20% sample condition (monkey 2, 20% sample: slope, $t(1,8) = 43.3$, $q < .001$, PSE, $t(1,8) = 13.2$, $q < .001$; 30% sample: slope, $t(1,8) = 4.4$, $q = .0696$, PSE, $t(1,8) = 18.6$, $q = .919$; 40% sample: slope, $t(1,8) = 2.4$, $q = .159$, PSE, $t(1,8) = 18.4$, $q < .001$). For this subject, despite the fact that performance had improved within the period of flanker training itself, the addition of flankers had caused such a drastic drop in performance, that any gains made during flanker training merely contributed to a recovery in performance to pre-flanker levels. Ultimately, this subject failed to improve beyond the peak levels that had been reached prior to flanker training.

3.3.2 Effects of adding flanker stimuli

3.3.2.1 Perceptual learning for individual test contrast conditions

To examine how learning rates differed between test contrast conditions, performance was plotted separately for each test contrast (refer to Figure 56).

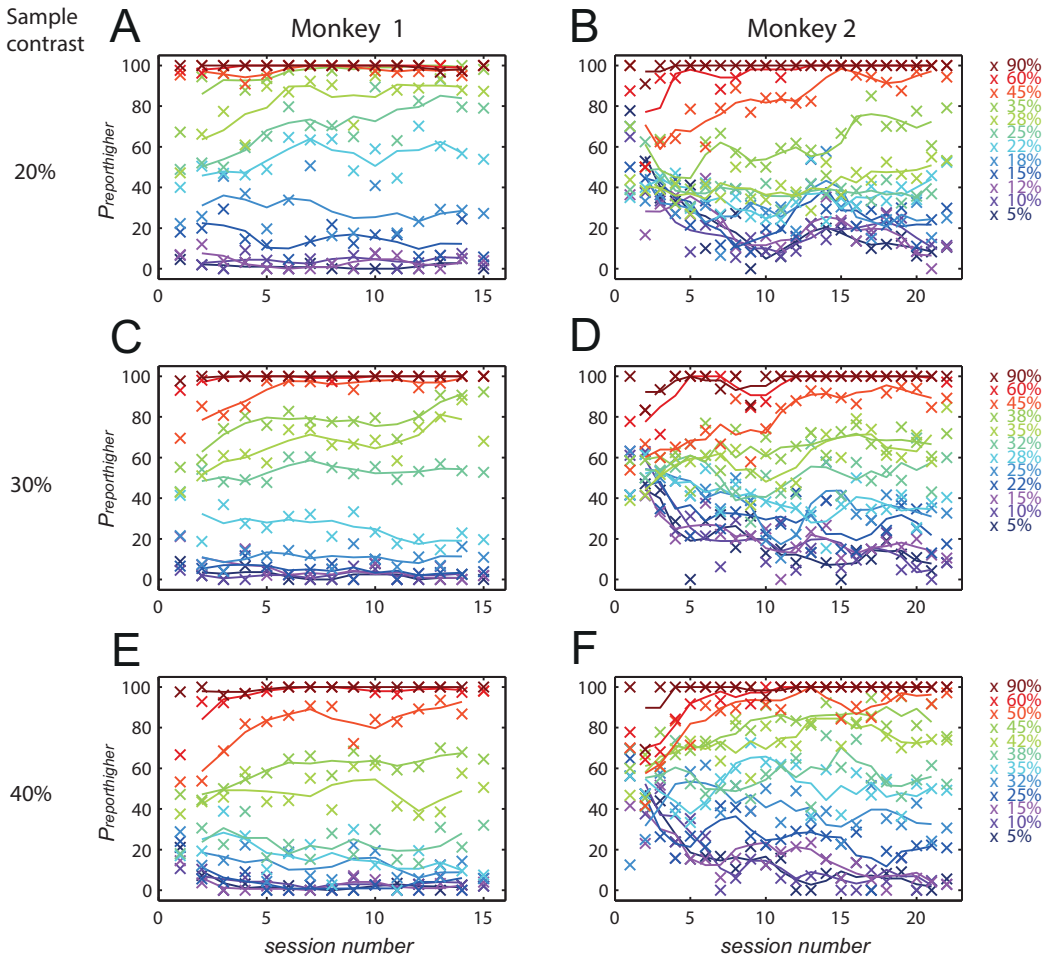


Figure 56. Proportion of trials during which the contrast of the test stimulus was reported to be higher than that of the sample, for each test contrast condition (coded by colour), plotted against session number, during flanker training. Left column: monkey 1; right column: monkey 2. A & B: 20% contrast sample; C & D: 30% contrast sample; E & F: 40% contrast sample. 'X' markers correspond to raw data points, while lines represent the running average, calculated across three consecutive sessions at a time.

The greater the difference between sample and test contrasts, the better the subjects' performance, and the faster an asymptotic level of performance was reached. In monkey 2, a particularly marked asymmetry in learning emerged, in which

improvements for low test contrasts lagged behind those for high test contrasts (indicated by the shallower slope of the best fit curve for low test contrasts than for high ones). Incidentally, this pattern was similar to that seen in this subject during training on the non-roving task, when stimuli were located in the V4 location (compare the above figure with Figure 7 on page 29).

3.3.3 Psychometric thresholds during the flanker task

A Spearman's rank correlation analysis was carried out between threshold and session number, to test for changes in threshold over time. Significant decreases were observed over the course of training for all upper threshold values, as well as for the majority of lower thresholds (Table 29). These widespread improvements matched those observed in the other parameters of performance.

Statistic	*df*	*r*	*q*	*df*	*r*	*q*
	Monkey 1			Monkey 2		
			20%			
C_L	13	-.304	.271	20	-.408	.0591
C_H	13	-.682	.00653*	20	-.673	< .001*
			30%			
C_L	13	-.609	.0159*	20	-.672	< .001*
C_H	13	-.764	.00139*	20	-.810	< .001*
			40%			
C_L	13	-.529	.0454	20	-.889	< .001*
C_H	13	-.836	< .001*	20	-.692	< .001*

* $q < \alpha$

Table 29. Changes in psychometric thresholds during the roving task. FDR correction for multiple comparisons, $\alpha = .05/12 \times 9 = .0375$.

3.3.4 Reaction times

For each session, mean RTs were calculated separately for correct and incorrect trials, across all 12 test contrast conditions. No significant changes in RT were observed during this stage of training (Table 30, Pearson's correlation coefficient, FDR correction for multiple comparisons, $\alpha = .05/12 = .0042$).

	Monkey 1		Monkey 2	
Statistic	*r*	*q*	*r*	*q*
	20% contrast			
Correct	-.110	.695	.282	.203
Error	-.119	.673	.511	.0152
	30% contrast			
Correct	-.266	.338	.431	.0451
Error	-.0846	.764	.278	.211
	40% contrast			
Correct	-.135	.631	.168	.455
Error	-.135	.632	.239	.284

Table 30. Pearson's correlation coefficients and *q*-values for correlations between mean RT and session number during training on the roving task with flankers (FDR correction, $\alpha = .05/12 = .0042$).

3.3.5 Discussion of behavioural results from the flanker task

The proportion of correct trials and the slope of the psychometric function increased significantly for both subjects, across all sample contrasts. Further improvements were observed in the slope and PSE for certain sample contrast conditions, depending on the subject. On the whole, substantial gains were made by both subjects over this period of training, yielding much better performance at the end of flanker training, compared to the beginning.

However, when performance levels during pre-flanker sessions were taken into account, this revealed a striking divergence in performance between the two subjects upon the addition of flankers- for monkey 1, flankers induced a brief worsening of performance, followed by a rapid return to pre-flanker levels, and a subsequent surge in performance above that seen in the absence of flankers. For monkey 2, on the other hand, the addition of flankers triggered a plunge in his performance which, throughout the course of flanker training, never completely recovered to pre-flanker levels.

The use of a fixed flanker contrast raised the concern (described by Yu et al. (2004)) that observers might form a stimulus template at each sample contrast, based on an observation of the difference in contrast between the flankers and the central stimuli. Yu et al. addressed this by carrying out two versions of the task- one in which flanker contrasts were 'jittered' randomly from trial to trial, but remained the same during both

stimulus presentation intervals per trial; and one in which the flanker contrast was fixed at 40%. After analysing their data, they felt that this precaution had been unnecessary as no significant difference in results was found between the two versions of the task.

This concern, however, still exists for our paradigm as training in our study was carried out for a much longer period than in the human study. Subjects might have been able to build up 'difference templates' that captured the differences in contrast between the flankers and central stimuli, rather than the absolute value of sample and test stimuli (for a detailed description of how the monkeys may have used different strategies to carry out the task, see the section titled '*Possible differences in task strategy*,' page 199).

However, we did not view this as an unwelcome possibility, as such a strategy, while somewhat removed from the original requirements of the task (i.e. to perform a comparison between stimuli from separate time intervals), is biologically plausible and could still yield informative results. We could theoretically have run subjects on an additional training paradigm using jittered flanker contrasts; however, in practice, there was no way to explicitly instruct our monkeys to make their comparisons between the central stimuli, rather than between flanker stimuli, and this feature would have made their task exceedingly difficult. Furthermore, our subjects had already been trained using flankers of fixed contrast; previous studies have shown that prior exposure to task-related stimuli is able to enhance subsequent task performance, even when the initial period of exposure does not involve conscious attention to the stimuli (Watanabe, 2001), or involves different task requirements (Xiao et al., 2008). Thus, any attempt to introduce a period of training using jittered flanker contrasts would have been confounded by the prior experience of our subjects and would quite likely have made our results difficult to interpret.

3.4 Flanker task neuronal results

3.4.1 Changes in the CRF during training on the flanker task

3.4.1.1 Individual channel results

Changes were assessed based on four measures of the contrast response function, obtained during presentations of the test stimulus. Significant changes were seen on a very small number of channels (Table 31). Thus, the marked improvement observed at the behavioural level was not closely matched by changes at the level of individual neuronal CRFs in V1.

		Monkey 1			Monkey 2		
Sample contrast (%)		20	30	40	20	30	40
Slope versus session	Steeper	0	0	0	0	1	0
	Shallower	1	0	1	1	1	2
C_{50} versus session	Towards C_{sample}	2	2	1	0	0	0
	Away from C_{sample}	0	0	0	2	1	1
Minimum versus session	Increase	0	0	0	0	0	0
	Decrease	0	0	0	0	0	0
Maximum versus session	Increase	0	1	0	1	0	0
	Decrease	1	1	1	1	1	1

Table 31. Number of channels with significant changes in each parameter of the contrast response function, during training with flanker stimuli (monkey 1: $N = 23$; monkey 2: $N = 25$, FDR correction for multiple parameters).

3.4.1.2 Population results

To identify changes in the CRF during training with flanker stimuli, population CRFs were plotted for each session, and the four parameters obtained from the fitted curves were plotted against session number (Figure 57).

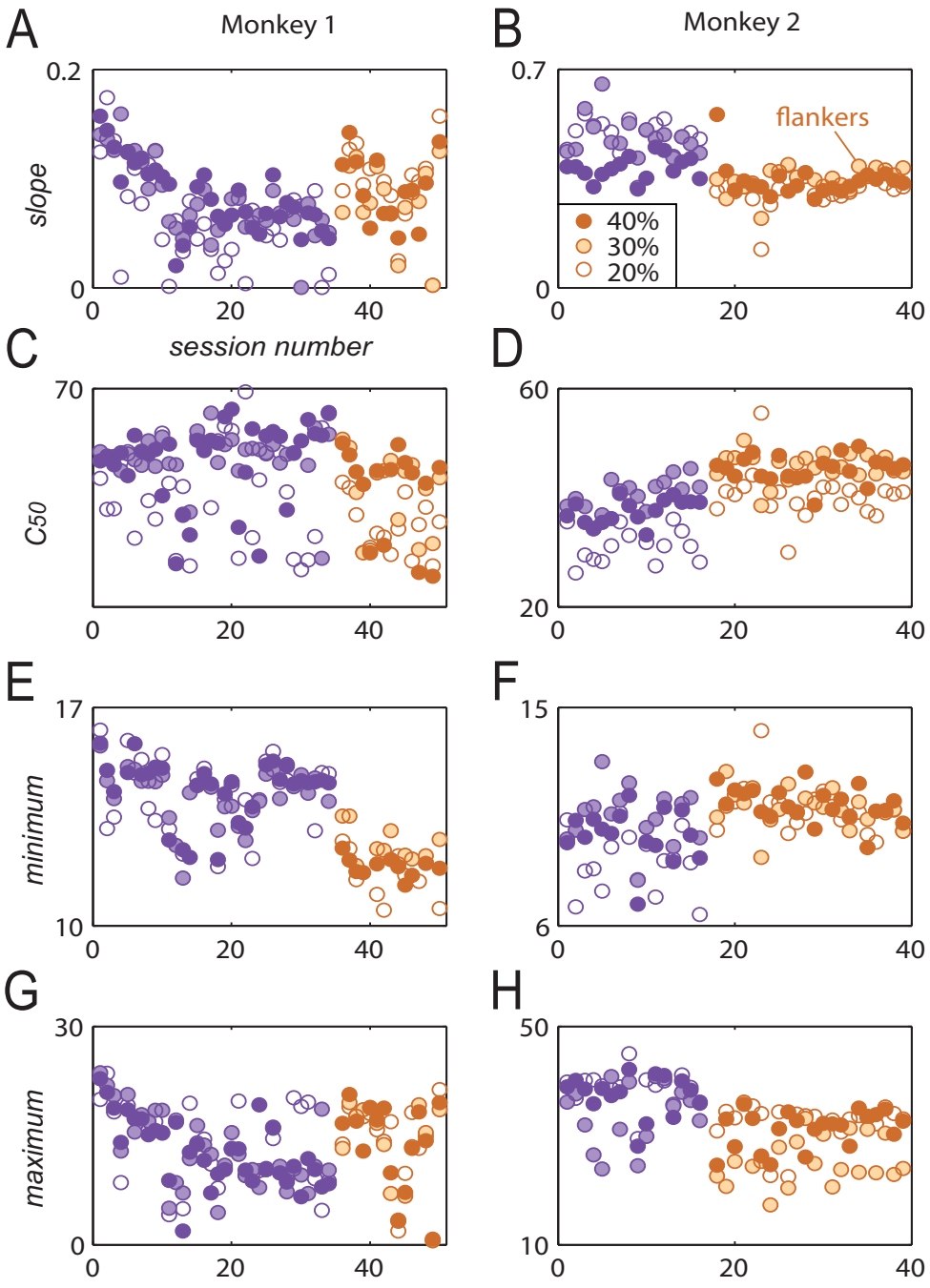


Figure 57. Parameter values of the population CRFs with time, during roving training, before (purple) and after the addition of flankers (orange). Note that purple markers present the same results as those shown in Figure 49, for comparison (page 137). Left column: monkey 1; right column: monkey 2. A & B: slope, C & D: C_{50}, E & F: minimum value; G & H: maximum value. Unfilled markers: 20% sample; medium purple/orange: 30%; dark purple/orange: 40%. During training with flanker stimuli, a shift in the C_{50} was observed for monkey 1, when the sample contrast was 40% (see Table 32).

During training with flankers, no significant changes in the CRF parameters were observed. Thus, changes in the V1 CRF did not appear to be able to account for the widespread improvements in performance seen for all three sample contrasts in both subjects at the behavioural level; note though, that in monkey 1, this may be due in part to the relatively small number of sessions that were conducted in the presence of flankers. When data were combined across the three sample contrast conditions, and a *robustfit* was performed to identify changes in each parameter with training, the minima showed a decrease for monkey 2 ($r(64) = -.045$, $q < .001$).

	Spearman's rank correlation								
Sample contrast (%)	20			30			40		
Statistic	*df*	*r*	*q*	*df*	*r*	*q*	*df*	*r*	*q*
	Flankers								
	Monkey 1								
Slope	13	-.125	.658	13	-.043	.883	13	-.314	.254
C_{50}	13	-.296	.283	13	-.250	.368	13	-.336	.221
Min	13	-.036	.903	13	-.282	.307	13	-.396	.145
Max	13	-.225	.419	13	-.254	.361	13	-.261	.347
	Monkey 2								
Slope	20	.294	.183	20	.478	.0257	20	.025	.912
C_{50}	20	-.149	.508	20	-.025	.912	20	.119	.596
Min	20	-.483	.0242	20	-.326	.139	20	-.416	.0552
Max	20	-.124	.582	20	.095	.672	20	.281	.205

* $q < \alpha$

Table 32. A Spearman's rank correlation analysis was carried out to identify changes in the slope, C_{50}, and minimum and maximum values of the CRF, during training with flanker stimuli. No significant changes were seen for individual sample contrast conditions, though a decrease in the minima was seen for monkey two when data were pooled across conditions (see text for details, FDR correction: $\alpha = .05/24 = .00208$).

While few signs of change were observed within the period of flanker training itself, a visual inspection revealed that the insertion of flankers had triggered an abrupt change in the CRF, from pre-flanker to flanker training (Figure 57). A two-way ANOVA was carried out for each parameter, comparing responses from the last 30% of pre-flanker sessions with those from the first 30% of flanker sessions. The two factors were the presence of flankers (absent or present) and the sample contrast (20, 30 or

40%). A significant main effect of flanker presence was observed in both subjects (Table 33).

	Monkey 1			Monkey 2		
	df	*F*	*q*	*df*	*F*	*q*
Slope	1,36	30.6	< .001*	1,24	24.3	< .001*
C_{50}	1,36	2.5	.125	1,24	22.8	< .001*
Minimum	1,36	6.0	.0190*	1,24	19.9	< .001*
Maximum	1,36	20.2	< .001*	1,24	25.5	< .001*

* $q < \alpha$

Table 33. A comparison of CRF parameters between the last third of pre-flanker training and the first third of flanker training revealed that the addition of flankers had brought about a significant change across numerous parameters in both monkeys (FDR correction: $\alpha = .05/8 \times 7 = .0438$).

In monkey 1, the addition of flankers was accompanied by a significant increase in the slope of the CRF. The maximum response increased upon addition of flankers, while the minimum response decreased. This corresponded to an increase in the range of spiking activity.

In monkey 2, the opposite pattern was seen. The slope decreased upon the addition of flankers; and the minimum increased while the maximum decreased, corresponding to a narrowing of the range of the CRF. In addition, the C_{50} was higher (i.e. further away from the sample contrasts) during flanker training, compared to during pre-flanker training.

Overall, this showed that neurometric performance (in terms of contrast sensitivity) improved in monkey 1 upon the addition of flankers, but worsened in monkey 2. The differences in the direction of modulation of firing rates between the two monkeys mirrored that seen in the behavioural data, in which monkey 1's performance was rapidly boosted by the addition of flankers (Figure 54), whereas monkey 2's performance deteriorated (Figure 55).

3.4.2 Changes in PROBMAT during training with flanker stimuli

3.4.2.1 Individual channel results

Neuronal data from individual channels were monitored for changes in the PROBMAT function over the course of training with roving samples. On the whole, few channels showed significant changes (Table 34). This matched the results seen for the CRF, in which behaviourally-linked changes were scarce at the level of individual channels.

		Monkey 1			Monkey 2		
		20	30	40	20	30	40
Slope versus session	Steeper	0	0	0	0	0	0
	Shallower	0	1	0	0	0	1
PNE versus session	Towards C_{sample}	0	0	0	0	1	1
	Away from C_{sample}	0	0	0	0	0	0
Minimum versus session	Increase	0	0	0	1	0	2
	Decrease	0	0	0	1	3	1
Maximum versus session	Increase	0	0	0	0	0	1
	Decrease	0	0	0	0	1	0

Table 34. Number of channels with significant changes in each parameter of the PROBMAT-versus-contrast function, during training with roving sample stimuli (monkey 1: $N = 23$; monkey 2: $N = 25$, FDR correction for multiple parameters).

3.4.2.2 Population results

As with the results from individual channels, few significant changes were seen within the period of flanker training (Table 35). However, a non-significant shift in the PNE towards 40% was seen for monkey 1 when a 40% sample contrast was presented (dark orange markers in Figure 58), mirroring the shift in the C_{50} observed for the CRF. This was partially in keeping with the behavioural improvement observed in monkey 1 (though one would also have expected to see changes for the 20% and 30% sample contrasts, if neurometric and psychometric performance were closely matched).

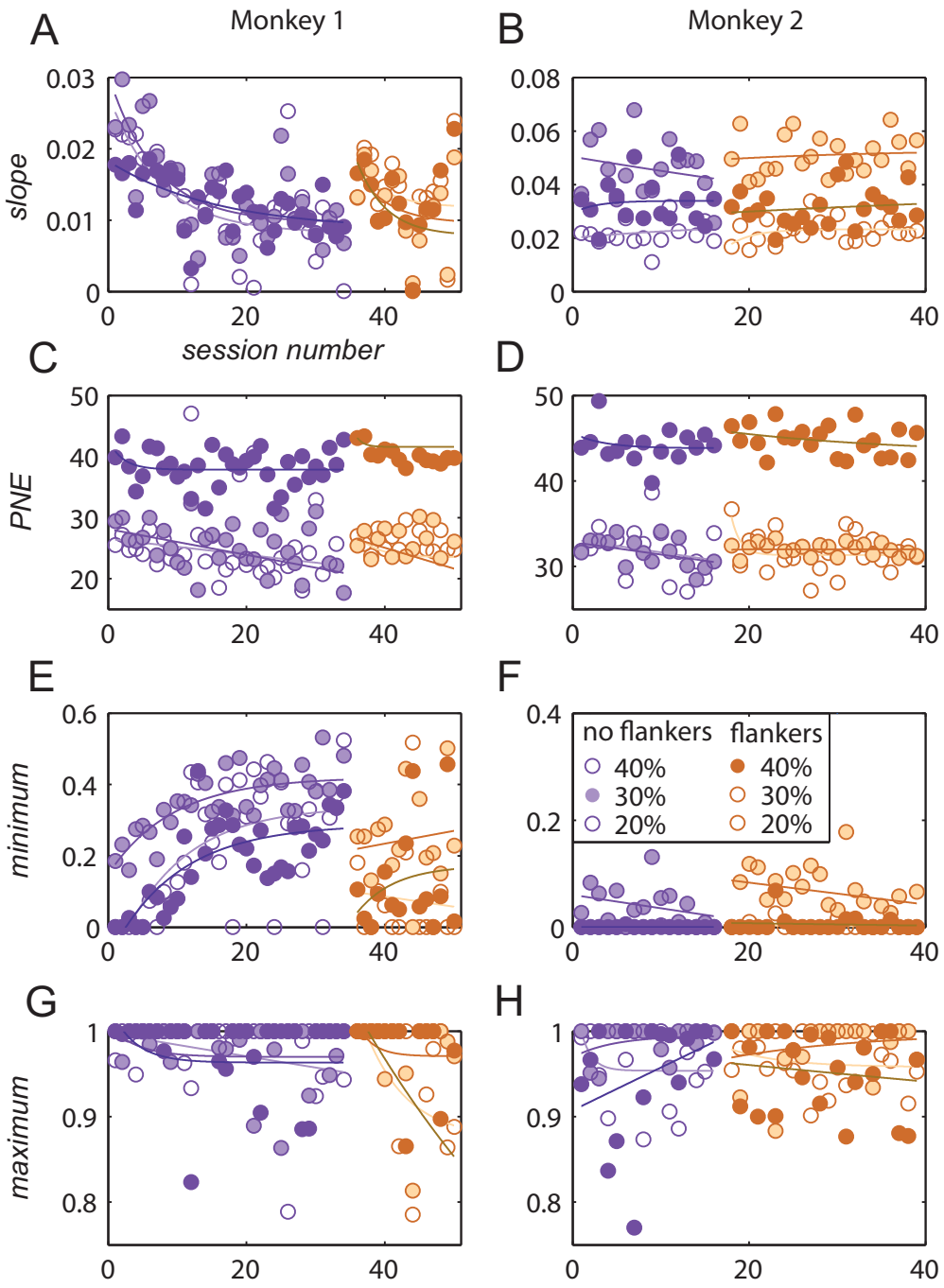


Figure 58. Parameter values of the population PROBMAT curve as subjects were trained on a roving stimulus task- initially in the absence of flankers (results from Figure 50 are marked here in purple for comparison), and then in the presence of flankers (orange). Left column: monkey 1; right column: monkey 2. A & B: slope; C & D: PNE; E & F: minimum value; G & H: maximum value. Unfilled markers: 20%; medium purple/orange: 30%; dark purple/orange: 40%. During training with flanker stimuli, no changes were observed in either subject (see Table 35).

Sample contrast (%)	Spearman's rank correlation								
	20			30			40		
Statistic	*df*	*r*	*q*	*df*	*r*	*q*	*df*	*r*	*q*
	Flankers								
	Monkey 1								
Slope	13	-.086	.763	13	-.282	.307	13	-.182	.515
PNE	13	-.204	.466	13	-.093	.743	13	-.604	.0195
Min	13	-.125	.658	13	-.007	.985	13	.046	.873
Max	13	-.556	.0314	13	-.318	.248	13	-.529	.0454
	Monkey 2								
Slope	20	.173	.439	20	.171	.445	20	.141	.531
PNE	20	-.091	.687	20	-.240	.282	20	-.242	.276
Min	20	-.197	.378	20	-.396	.0693	20	-.013	.956
Max	20	-.190	.395	20	.123	.586	20	-.124	.582

Table 35. A Spearman's rank correlation analysis was performed to identify changes in the slope, PNE, and minimum and maximum values of the neurometric function, during training on the roving task with flanker stimuli. No significant changes were seen for either monkey (FDR correction, $\alpha = .05/24\times6 = .0125$).

As with the CRF analysis, a two-way ANOVA was carried out to compare responses from the last 30% and the first 30% of pre-flanker and flanker sessions, respectively, with the presence of flankers and sample contrast as factors.

In monkey 1, the slope of the PROBMAT function was significantly higher, and the minimum was significantly lower, when flankers were added (slope: $F(1,36) = 8.46$, $q = .0062$; minimum: $F(1,36) = 41.6$, $q < .001$). A non-significant trend for an increase in the maximum was seen ($F(1,36) = 4.0$, $q = .0527$). These effects all corresponded to increases in discriminability.

In monkey 2, the PNE increased away from the sample contrasts ($F(1,24) = 7.83$, $q = .0100$), and the minimum became significantly higher ($F(1,24) = 7.9$, $q = .0096$), reflecting a worsening in neurometric performance.

In summary, when an examination of the data was restricted to the flanker training period alone, only a few changes were observed at the neuronal level. These changes did not appear sufficient to account for the widespread improvements observed in behavioural performance on the CD task in the presence of flankers. However, when

data were analysed across both pre-flanker and flanker training periods, while taking the performance of our subjects into account, a consistent pattern emerged- good task performance was associated with higher slopes and wider ranges of the CRF and PROBMAT functions, and shifts in the C_{50} and PNE towards the sample contrast, while poor task performance was associated with lower slopes and reduced ranges of the CRF and PROBMAT functions, and C_{50} and PNE values that were further removed from the sample contrast.

3.4.3 Neurometric thresholds during the flanker task

Thresholds were monitored over time for training-induced changes, in the presence of flanker stimuli (Table 36 and Figure 59). Upon addition of flankers, decreases had been seen in the psychometric thresholds of both subjects in 9/12 comparisons, predominantly when the test stimulus was of higher contrast than the sample (Table 29). However, no changes were seen in the neurometric thresholds.

Statistic	*df*	*r*	*q*	*df*	*r*	*q*
	Monkey 1			Monkey 2		
			20%			
C_L	13	-.065	.817	20	-.099	.661
C_H	13	.337	.220	20	-.138	.538
			30%			
C_L	13	.309	.262	20	-.094	.676
C_H	13	.313	.256	20	-.195	.383
			40%			
C_L	13	.258	.353	20	.105	.642
C_H	13	.168	.548	20	-.031	.892

$q < \alpha$

Table 36. No changes in neurometric thresholds were observed during the flanker task (Spearman's rank correlation). FDR correction for multiple comparisons, $\alpha = .05/12 = .0167$.

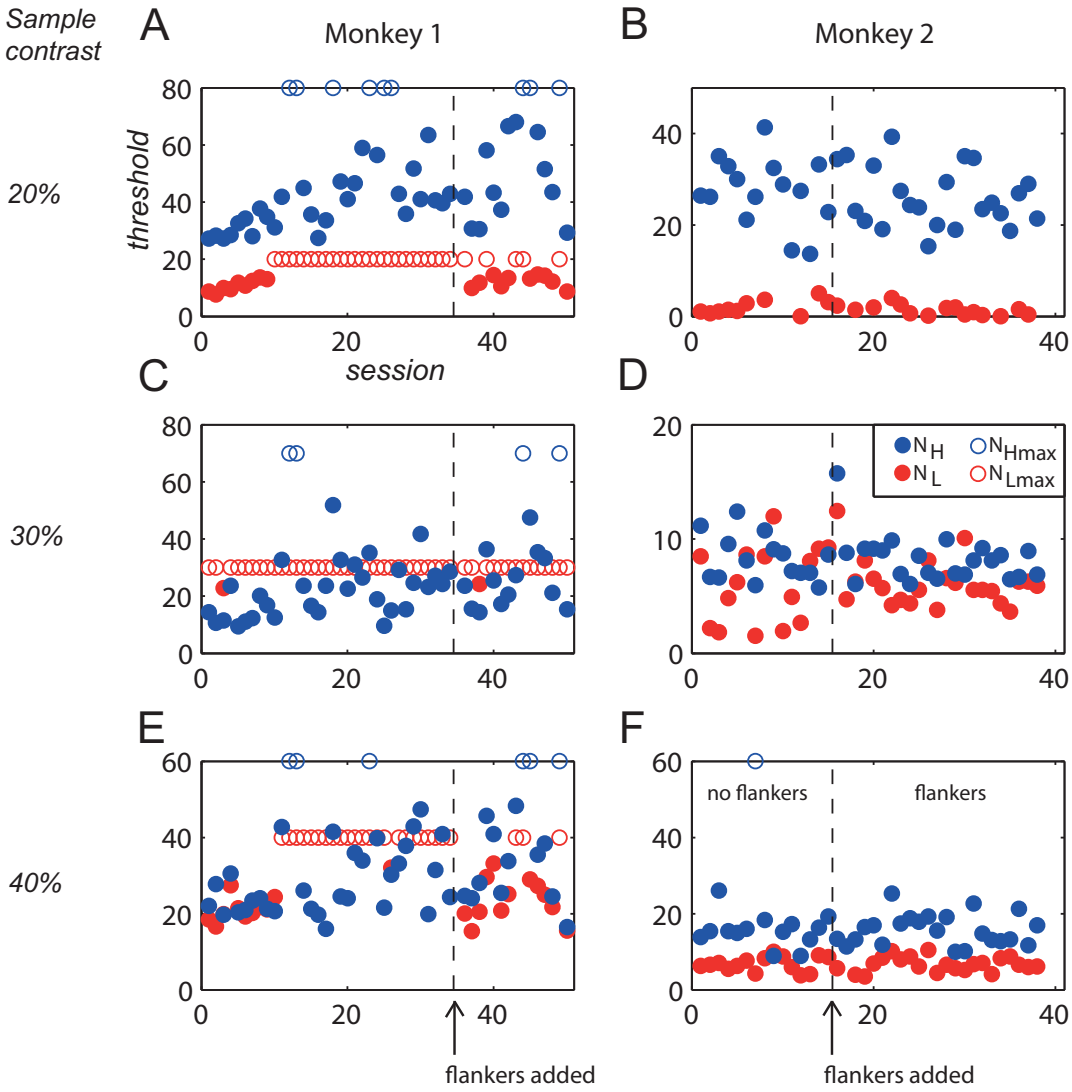


Figure 59. Neurometric thresholds (filled markers), plotted as a function of time. Unfilled markers indicate sessions where thresholds could not be obtained. Subjects were trained on a roving stimulus task, initially in the absence of flankers (results from Figure 51 are repeated here for comparison), and then, after the addition of flanker stimuli (vertical black line, annotated with an arrow), in the presence of flankers. Left column: monkey 1; right column: monkey 2. A & B: 20% sample; C & D: 30% sample; E & F: 40% sample. Red markers: N_L conditions (the test contrast was lower than that of the sample); blue markers: N_H conditions (the test contrast was higher than that of the sample). No significant decreases in threshold value were observed (refer to Table 36 for results from the correlation analysis).

3.4.4 Variability of the visual response during training with flankers

Upon the addition of flankers, changes in the *FF* were inconsistent between the two monkeys. At the level of individual channels, the FF increased for all channels in

monkey 1 (20% sample: 23/23 channels, 30%: 23/23 channels, 40% sample: 23/23 channels) and decreased for the majority of channels in monkey 2 (20% sample: 18/25 channels, 17/18 decreases, 1/18 increase; 30%: 21/25 channels, 20/21 decreases, 1/21 increase; 40% sample: 18/25 channels, 18/18 decreases, 0/18 increases).

At the population level, following the addition of flanker stimuli, the *FF* increased during all three sample contrast conditions in monkey 1, whereas it decreased for all three sample conditions in monkey 2 (Table 37), in keeping with the observations made with individual channel data.

	Subject 1			Subject 2		
Sample contrast (%)	*df*	*F*	*q*	*df*	*F*	*q*
	flankers					
20	1,2184	276.5	< .001*	1,3576	68.1	< .001*
30	1,2184	280.4	< .001*	1,3576	93.5	< .001*
40	1,2184	278.3	< .001*	1,3576	102.9	< .001*

* $q < \alpha$

Table 37. Results from a two-factor ANOVA, comparing trial-wise spike variability between early and late flanker sessions. Significant changes in the Fano factor occurred in all cases when flankers were present (FDR correction for multiple comparisons, $\alpha = .05/6\times6 = .05$).

3.4.5 Discussion of neuronal results from the flanker task

In the previous chapter on pre-flanker roving training, neuronal changes were inconsistent between the two monkeys. The subsequent introduction of flankers amplified differences in their performance, at both the behavioural and the neuronal level. While the lack of uniformity across subjects makes it difficult to make broad conceptual generalisations to contrast discrimination tasks as a whole, it does offer a realistic view of the situation seen in human studies (i.e. effects differ, depending on the individuals performing the task). Adini et al. (2004) found some improvement during roving training with flankers when a blocked paradigm was used, and somewhat less improvement when a MBT method was used, while Yu et al. (2004) found only scattered improvement for a roving task in the absence of flankers, and none whatsoever in the presence of flankers.

Importantly, the differences observed at the neuronal level in the current study may account for disparities in psychometric performance across subjects. Improvements at the behavioural level with the 40% sample were seen in monkey 1, and these changes were reflected in shifts in the C_{50} and the PNE towards 40%. Thus, V1 activity underwent discernible learning-induced modulation during training on a roving task with flankers.

While monkey 1 improved significantly on the flanker task, and modulations of spiking activity on his V1channels reflected his progress, monkey 2's performance languished in comparison to pre-flanker levels, and his neurometric performance showed no change during flanker training, beyond a narrowing in the range of the population CRF.

A comparison of activity levels during late sessions of pre-flanker training with early sessions of flanker training revealed several behaviourally-coupled changes: a widening in the range of spiking activity upon the addition of flankers occurred in monkey 1, and a shrinking occurred in monkey 2. Increases in discriminability were obtained for monkey 1, whereas decreases in the slopes of the CRF functions and shifts of the C_{50} and PNE away from the sample contrasts occurred for monkey 2. These patterns correlated well with the subjects' reaction to the new task, and the differences in the direction of neurometric function modulation between the two monkeys closely mirrored those seen in their behavioural performance.

Similarly, the direction of changes in the variability of V1 spiking activity paralleled those seen on the other measures of neurometric and psychometric performance. Improvements were associated with increases in the *FF* (as seen in monkey 1), whereas deteriorations were associated with decreases in the *FF*. In the previous chapter, the *FF* decreased in monkey 1 over the course of training, when improvement on the flankerless roving task was relatively modest. In monkey 2, an increase in the *FF* occurred for the 20% sample condition, i.e. the condition for which improvements in performance were seen. When the results of the previous chapter were examined in isolation, the findings were hard to interpret and the directions of effects were at odds between the two monkeys. Placed in context with the findings from the current chapter on flanker training, however, clear trends emerge, in which increases in

the variability of spiking are correlated with gains in performance, whereas decreases in variability are associated with a worsening or stagnation in performance. These results are harder to reconcile with the trends seen during the non-roving task, however, as performance increased for both subjects and both recording sites, whereas the *FF* changed in different directions, depending on the location and subject. One possible explanation for the decreases in FF seen for monkey 1 in V1 and for monkey 2 in V4 might be that the prolonged period of training encompassed initial, dramatic improvement, followed by a plateau in performance, thus obscuring the relationship between spiking variability and CD performance.

In summary, in Chapter 2, during training on a flankerless roving task, while behavioural improvement was limited, it was not entirely absent; however, changes at the neuronal level were scant and failed to satisfactorily account for the handful of improvements observed. In this chapter, the addition of flankers had the (unintended) effect of magnifying differences in behavioural performance between the two subjects, and brought performance-dependent differences in V1 activity to light. Modulations in this region were closely linked to perceptual ability, at first glance giving credence to the argument that this may have been due to the involvement of V1 in PL.

Nevertheless, several other explanations should be considered at this juncture. If the two subjects had performed the flanker task using different strategies, this may have influenced their deployment of attention in the presence of flankers, leading to attention- and strategy-driven modulations of the V1 response. For example, monkey 1 may have used the flankers as fixed markers of 30% contrast, and his enhanced task performance may have arisen through a comparison of flanker and central stimuli during the two stimulus presentation intervals (this strategy is elaborated upon in the section, ‘Possible differences in task strategy,’ page 199). Monkey 2, on the other hand, was presented with much smaller stimuli, at the V1 location, and this factor may have made it harder for the subject to perceive the flanker stimuli as being separate from the central stimulus. If so, then the contrast differences between flanking and central stimuli may have been harder to resolve, hence leading to worse performance upon the addition of flankers (this scenario is also presented in the section, ‘Possible differences in task strategy,’ page 199). The changes observed in behavioural and neuronal performance, between pre-flanker and flanker training, may thus not have been due strictly to

perceptual learning, but rather to an interaction between attention modulation and subject-specific approaches to the task. Within the flanker training period, no clear correspondence between behavioural and neuronal data occurred, further supporting the possibility that the changes between pre-flanker and flanker training might have been attention- and/or strategy-related.

3.5 Removal of flanker stimuli

Finally, we removed the flankers so that our subjects performed the roving task using isolated sample and test gratings, as was done before the introduction of flankers. The goal of this stage of training was to determine whether flanker-induced changes persisted under flankerless conditions.

3.5.1 Behavioural results

3.5.1.1 Subjects performed a roving task, post-flanker-training

Subjects performed the roving task with a flankerless grating stimulus for several sessions, to enable a comparison of performance between this 'post-flanker' stage with that of the 'pre-flanker' stage (monkey 1: 7 sessions, spanning a 1.5 weeks; monkey 2: 4 sessions, spanning 1 week).

A visual inspection of levels of performance upon removal of flankers revealed that performance returned to pre-flanker levels (green markers, Figure 60 and Figure 61).

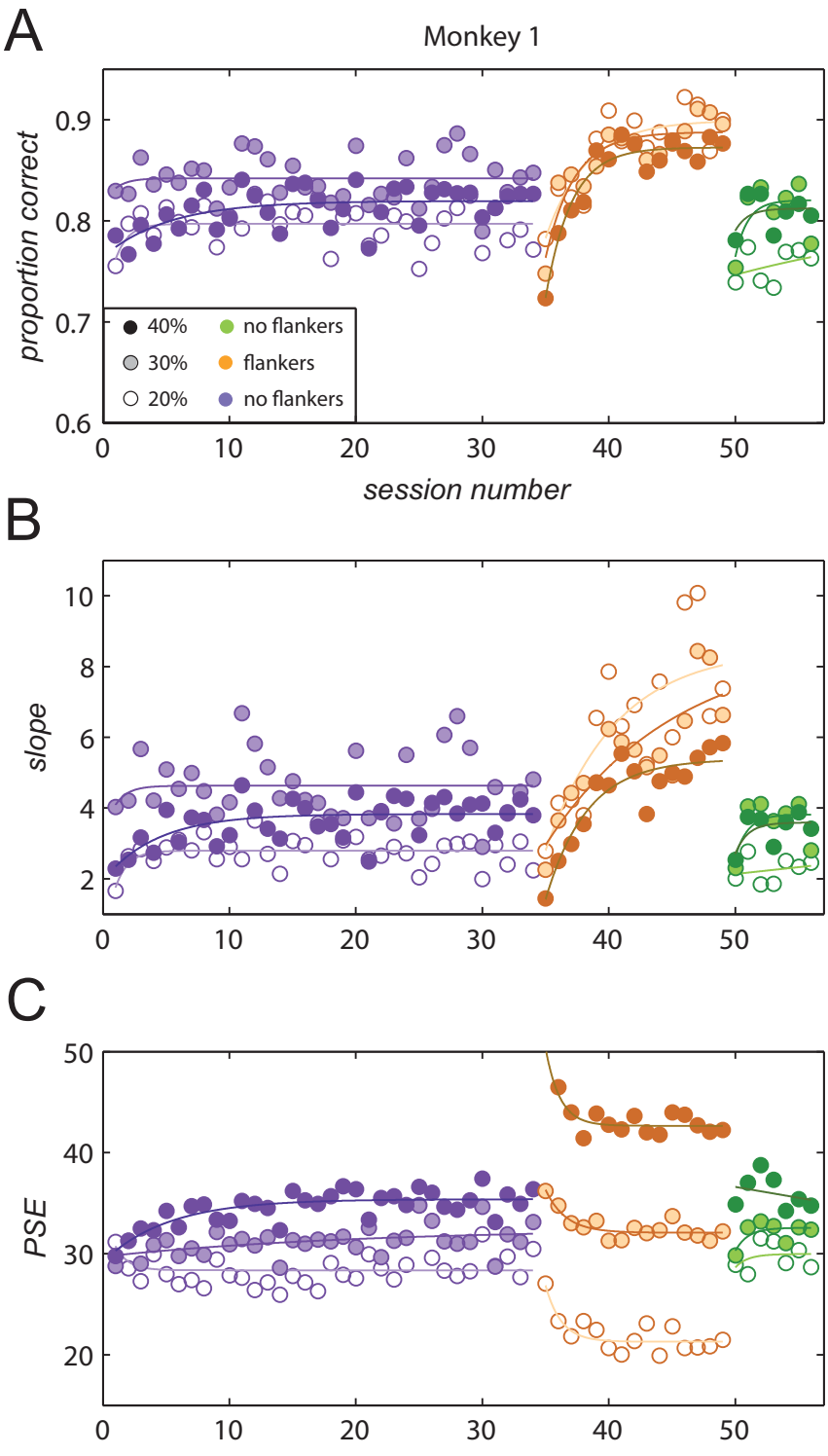


Figure 60. Overall performance of monkey 1 during the roving task. A: $P_{correct}$; B: slope of the psychometric function; C: PSE of the psychometric function. Purple data points: pre-flanker task; orange data points: flanker task; green data points: post-flanker task. Unfilled markers: 20% sample contrast conditions; medium-coloured filled markers: 30%; dark-coloured filled markers: 40%.

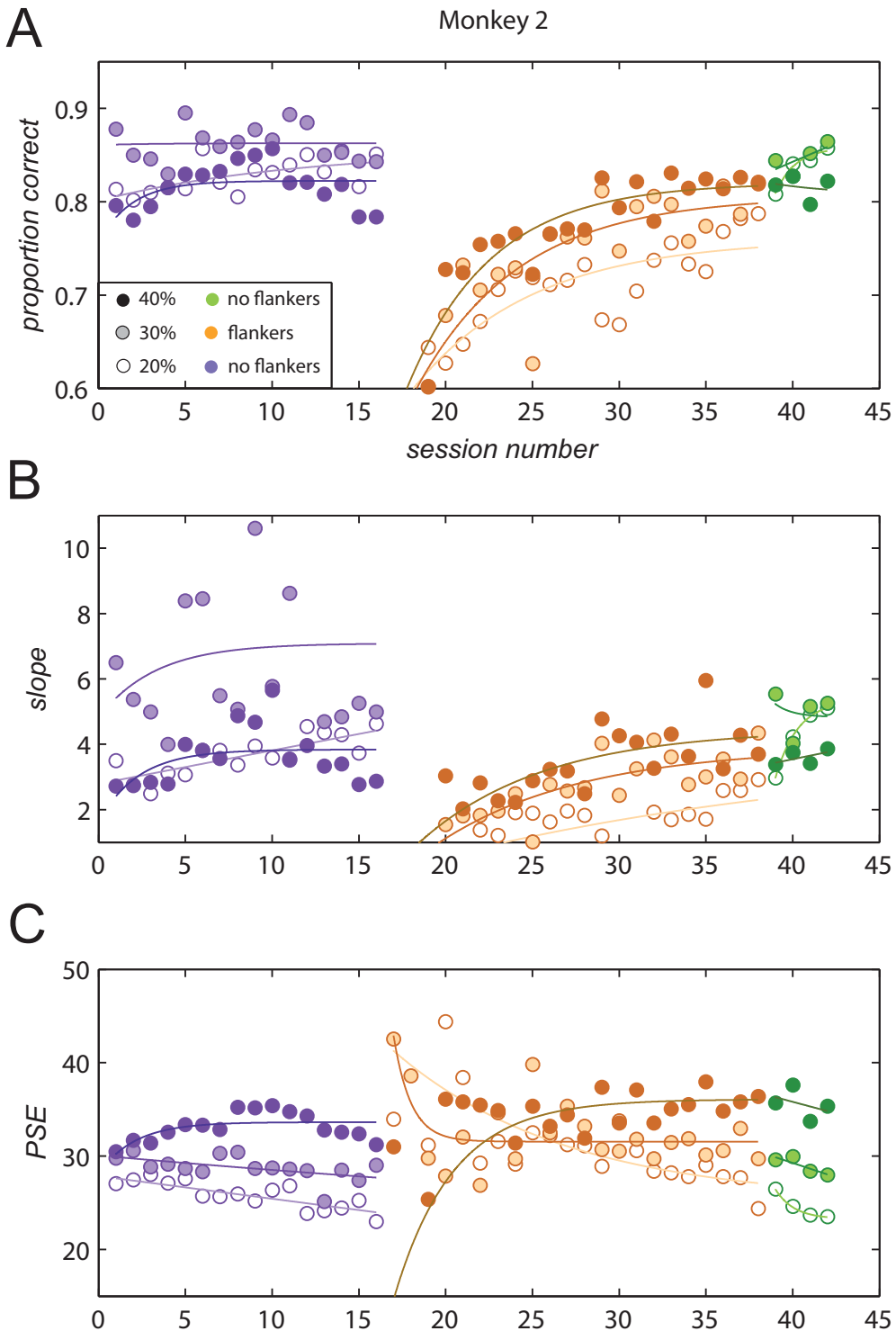


Figure 61. Overall performance of monkey 2 during the roving task. A: $P_{correct}$; B: slope of the psychometric function; C: PSE of the psychometric function. Purple data points: pre-flanker task; orange data points: flanker task; green data points: post-flanker task. Unfilled markers: 20% sample contrast conditions; filled, medium-coloured markers: 30%; filled, dark-coloured markers: 40%.

3.5.1.2 Effects of removing flanker stimuli on performance of the roving task

Levels of post-flanker performance were compared to those attained just prior to the introduction of flankers. This comparison allowed us to determine whether the changes seen during flanker training were contingent upon the presentation of flanker stimuli, or whether they would persist after the removal of flankers.

We anticipated that the subjects' performance during the first few sessions after flanker removal might be relatively poor, as they adjusted to the previous, flankerless version of the task. Thus, our analysis focused on data that was obtained from the last of these sessions.

For the most part, subjects' performance during this session (X_a) fell within the ranges of values seen during the late phase of the initial flankerless stage (Table 38). For monkey 1, the proportion of correct responses, the slope, the PSE, $RT_{correct}$ and RT_{error} lay within the ranges attained during the late phase of pre-flanker training for the 20% sample, while they were either within the ranges or slightly worse, for the 30% and 40% sample. For monkey 2, although $RT_{correct}$ and RT_{error} were worse during the last flankerless session, values of the slope fell within previous ranges for the 30% and 40% samples, while for the 20% sample, the proportion of correct responses, the slope, and the PSE were slightly better than before.

Thus, the monkeys' ability to discriminate contrast levels was largely comparable between sessions before and after training with flankers, indicating that any changes in performance that accompanied the addition of flankers were temporary and depended on the continued presentation of flankers.

	Monkey 1		Monkey 2	
	Late pre-flanker sessions, range $X_{min} – X_{max}$	Last post-flanker session, X_a	Late pre-flanker sessions, range $X_{min} – X_{max}$	Last post-flanker session, X_a
	20% sample			
$P_{correct}$ (%)	75.2 – 82.5	76.3	81.6 – 85.4	85.7
Slope	2.0 – 3.1	2.4	3.7 – 4.5	5.1
PSE	27.7 – 34.8	28.7	23.8 – 25.3	23.5
$RT_{correct}$	100.1 – 124.3	119.4	158.4 – 166.2	170
RT_{error}	110.6 – 148.5	136.9	166.5 – 174.7	179.1
	30% sample			
$P_{correct}$ (%)	78.9 – 88.6	77.7	84.3 – 88.4	86.4
Slope	2.9 – 6.6	2.8	4.7 – 14.1	5.3
PSE	28.7 – 34.7	32.4	25.1 – 28.5	28
$RT_{correct}$	103.0 – 118.9	120	157.3 – 167.4	170.2
RT_{error}	113.5 – 143.7	131.6	163.1 – 170.8	175.4
	40% sample			
$P_{correct}$ (%)	79.5 – 83.3	80.5	78.2 – 82.1	82.2
Slope	3.2 – 4.3	3.4	2.8 – 4.0	3.9
PSE	33.1 – 37.4	34.8	32.4 – 34.3	35.3
$RT_{correct}$	102.6 – 121.9	123.4	156.4 – 169.2	172.1
RT_{error}	91.7 – 136.9	115.3	154.6 – 166.4	169.7

Table 38. Comparison of subjects' performance in the absence of flankers, during post-flanker sessions, and during the end of pre-flanker sessions. $X_{min} – X_{max}$: Ranges of performance seen during late pre-flanker sessions, which took place before flankers were introduced. X_a: Performance recorded during the last session of post-flanker training, in which roving stimuli were presented, after the removal of flankers.

3.5.2 Discussion of post-flanker behavioural results

Changes in performance during training on the flanker task- whether in the form of improvements or deteriorations- did not persist in the absence of flankers. In monkey 1, performance on the flankerless task was even slightly worse after a period of flanker training. This result closely matches that reported by Yu et al. (2004), in which practise with flankers resulted in increases in contrast thresholds and partial reversals of pre-flanker improvements in performance.

If training with flankers had engaged exactly the same cognitive processes as those used in the absence of flankers, one would not expect to see a reversal in

performance after their removal. Based on our observations, the neuronal mechanisms used to perform the task in the absence of flankers appeared to be distinct from those used in the presence of flankers. The next section investigates whether underlying changes in spiking activity could account for the patterns observed.

3.5.3 Neuronal results

3.5.3.1 Changes in the CRF during training on the post-flanker task

Parameters of the CRF were plotted against time (Figure 62). A visual inspection revealed a pattern which closely matched that seen at the behavioural level: upon removal of flankers, parameter values returned to levels obtained during pre-flanker training.

Sessions were divided into three groups, depending on the training paradigm: the last 30% of pre-flanker sessions; all the flanker sessions; and all the post-flanker sessions. A two-factor ANOVA was carried out for each parameter, with training stage and sample contrast as factors.

A main effect of training paradigm was detected in the majority of instances. In most cases, values from the pre- and post-flanker stages were not significantly different from each other, but they were each significantly different from those seen during flanker training. In monkey 1, the C_{50} was significantly lower ($F(2,87) = 11.3$, $q < .001$) during flanker training, than that seen during either pre-flanker or post-flanker training. In monkey 2, the slope was significantly higher in the absence of flankers ($F(2,81) = 27.5$, $q < .001$), the C_{50} was significantly lower ($F(2,81) = 40.0$, $q < .001$), the minimum was also significantly lower ($F(2,81) = 32.3$, $q < .001$), and the maximum was significantly higher ($F(2,81) = 21.7$, $q < .001$). The only exception to this trend was for the slope of the CRF in monkey 1, where no difference was seen between flanker and post-flanker slopes, though they were each significantly higher than pre-flanker slopes ($F(2,87) = 7.76$, $q < .001$). An FDR correction was carried out for multiple comparisons, $\alpha = .05/8 \times 6 = .0375$.

These results confirmed the findings reported in the previous section: higher performance on the CD task (in the presence of flankers for monkey 1 but in the *absence* of flankers for monkey 2) was accompanied by steeper slopes, lower minima and higher maxima of the CRF, and shifts in the C_{50} towards the sample contrast, whereas poorer performance on the CD task (in the absence of flankers for monkey 1 but in the presence of flankers for monkey 2) was accompanied by shallower slopes,

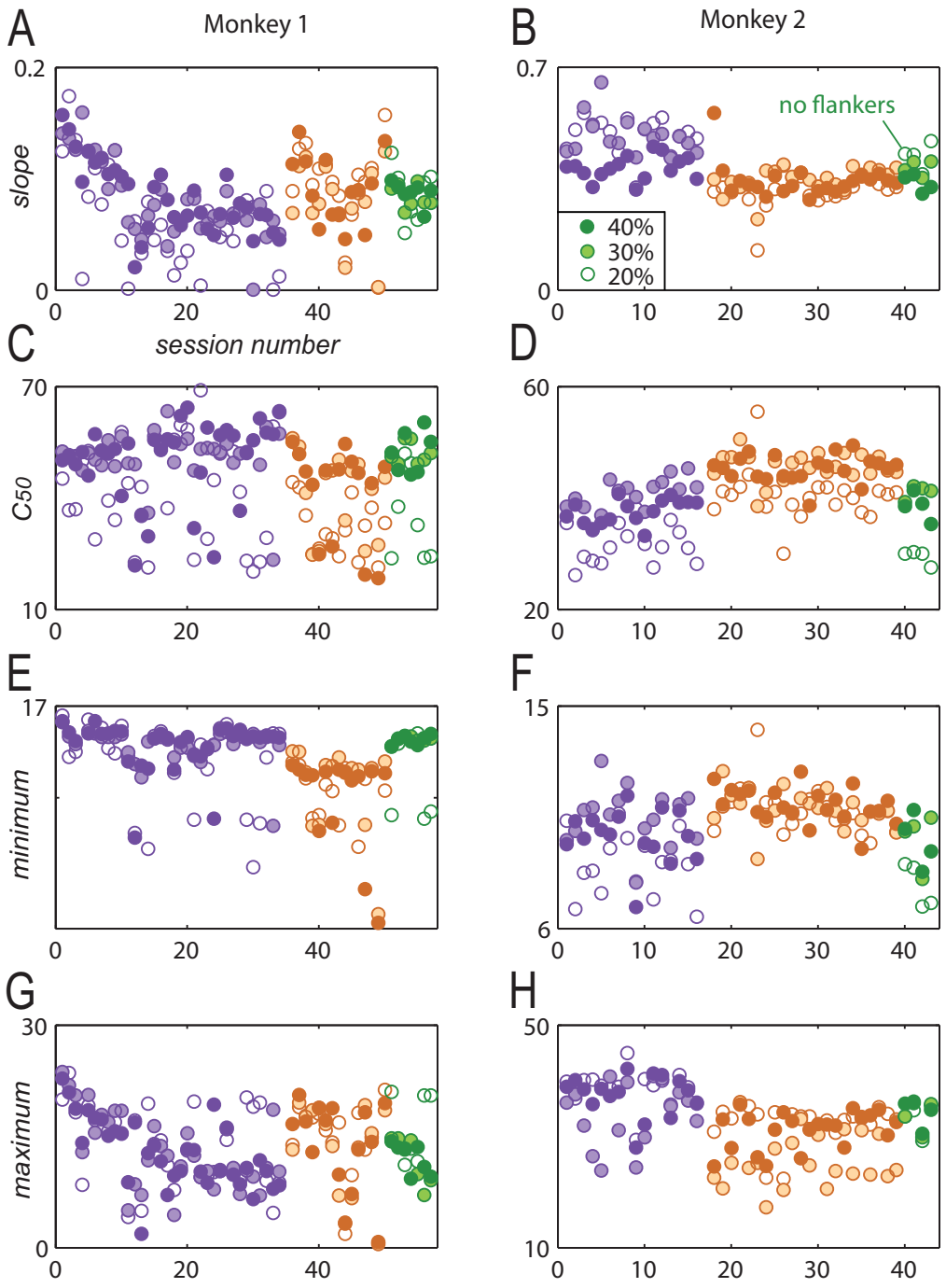


Figure 62. Parameter values of the population CRF with time, during roving training, after the removal of flankers (green). For comparison, purple and orange markers depict values during pre-flanker and flanker training, respectively (presented previously in Figure 57). Left column: monkey 1; right column: monkey 2. A & B: slope; C & D: C_{50}; E & F: minimum value; G & H: maximum value. Unfilled markers: 20% sample; medium purple/orange/green: 30%; dark purple/orange/green: 40%. In the absence of flanker stimuli, parameters of the CRF returned to the levels seen prior to the addition of flankers.

higher minima and lower maxima of the CRF, and shifts in the C_{50} away from the sample contrast.

3.5.3.2 Changes in PROBMAT during training on the post-flanker task

Parameters of the PROBMAT function were plotted against time (Figure 63). As with the analysis carried out on the CRF, sessions were divided into three groups: the last 30% of pre-flanker sessions; all the flanker sessions; and all the post-flanker sessions. A two-factor ANOVA was carried out for each parameter, with training stage and sample contrast as factors.

A significant main effect of training paradigm was observed in two of the eight comparisons. In monkey 1, the minimum of the PROBMAT function was significantly lower when flankers were present ($F(2,87) = 19.0$, $q < .001$). In monkey 2, the opposite effect was seen- the minimum of the PROBMAT function was significantly lower when flankers were *absent* ($F(2,87) = 6.3$, $q = .00276$, FDR correction, $\alpha = .05/8\times2 = .0125$). This corresponded to an enhancement in discriminability for monkey 1 and a worsening for monkey 2, during flanker training.

These changes in the PROBMAT function matched the patterns seen in the CRF, particularly in terms of the ranges of PROBMAT values. Ranges were narrower when performance was poor (during flankerless training for monkey 1 but during flanker training for monkey 2), and wider when performance was good (during flanker training for monkey 1 and during flankerless training for monkey 2).

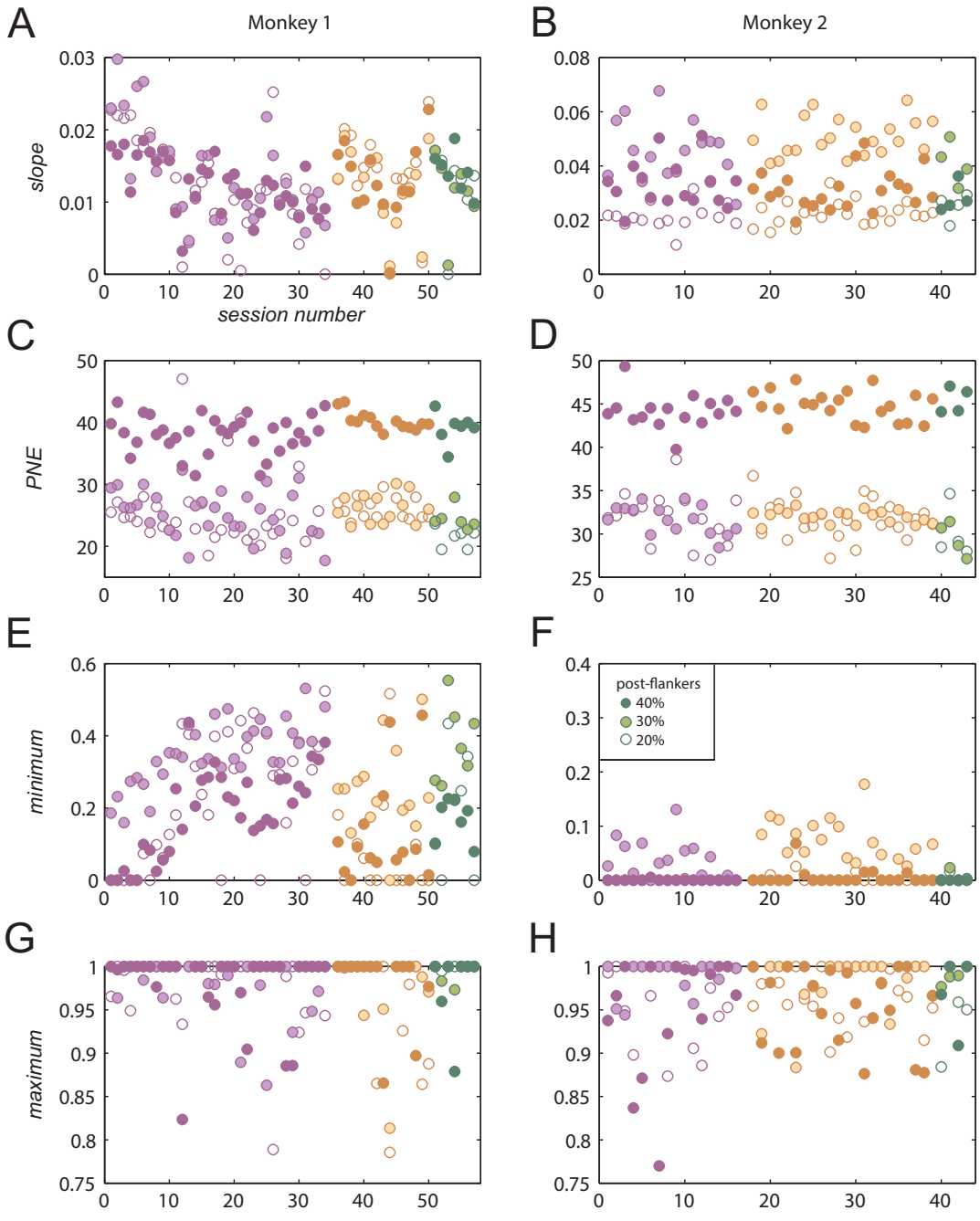


Figure 63. Parameter values of the population PROBMAT function after removal of flankers (green). Results from pre-flanker and flanker training in Figure 58 are marked here in purple and orange, respectively, for comparison. Left column: monkey 1; right column: monkey 2. A & B: slope; C & D: PNE; E & F: minimum value; G & H: maximum value. Unfilled markers: 20%; medium purple/orange: 30%; dark purple/orange: 40%.

3.5.4 Summary of all roving task results

When subjects were first presented with roving stimuli, they initially performed the contrast discrimination task based on a comparison with a sample of 30% contrast, regardless of the actual contrast of the sample. With regular training, improvement was possible, but it only occurred for a subset of conditions. Changes at the neuronal level in V1 were limited to overall declines in activity and did not appear to be closely linked to the changes in behaviour. Thus, the roving task was a challenging one and despite intensive, continuous practice, subjects' performance plateaued at levels that either matched or were only slightly better than those seen with the non-roving task.

The addition of flankers had an intriguing effect, which differed between the two subjects. In monkey 1, the presence of flankers boosted performance on the roving task, accompanied by a widening in the range of firing rates upon addition of flankers, and an increase in the slope of the neurometric function. Upon removal of flankers, performance dropped to the levels seen prior to flanker training, and V1 responses reverted to previous levels. In monkey 2, the addition of flankers was detrimental to performance, and brought about a decrease in CRF slope, a shift in the C_{50} and PNE away from the sample contrasts, and a narrower range of spiking activity and stimulus discriminability. Upon removal of flankers, performance levels were restored to their previous highs, and V1 responses returned to pre-flanker levels. The reversals induced by the removal of flankers strengthened the conclusion that V1 activity was closely linked to behavioural performance, and that modulations occurred not only during improvements in contrast discrimination, but also during deteriorations in task performance.

3.6 Correlations between psychometric and neurometric performance

An ultimate goal of our study was to examine whether correlations existed between spiking activity and the monkeys' performance on the CD task. We hypothesised that any correlations, if present, would occur regardless of the exact task

paradigm used (e.g. V4 or V1 location; non-roving or roving stimuli; absence or presence of flankers).

Thus, a correlation analysis was carried out between parameters of neurometric and psychometric performance, using data that were pooled across multiple CD task paradigms. First, neuronal data were *z*-scored within each task paradigm (to eliminate effects of between-area differences in spiking activity). *Z*-scored data were then combined across all the training periods. For each parameter of interest in the CRF and PROBMAT neurometric function, the values derived from the neuronal data were plotted against the proportion of correct trials and the slope of the psychometric function, providing an overview of neuronal versus psychophysical performance (Figure 64 and Figure 65). This included data across both V4 and V1 recording sites.

Comparisons were made between the following (*z*-scored) parameters for each of the monkeys:

1. CRF slope and psychometric slope
2. CRF slope and $P_{correct}$
3. C_{50} and PSE
4. PROBMAT slope and psychometric slope
5. PROBMAT slope and $P_{correct}$
6. PNE and PSE

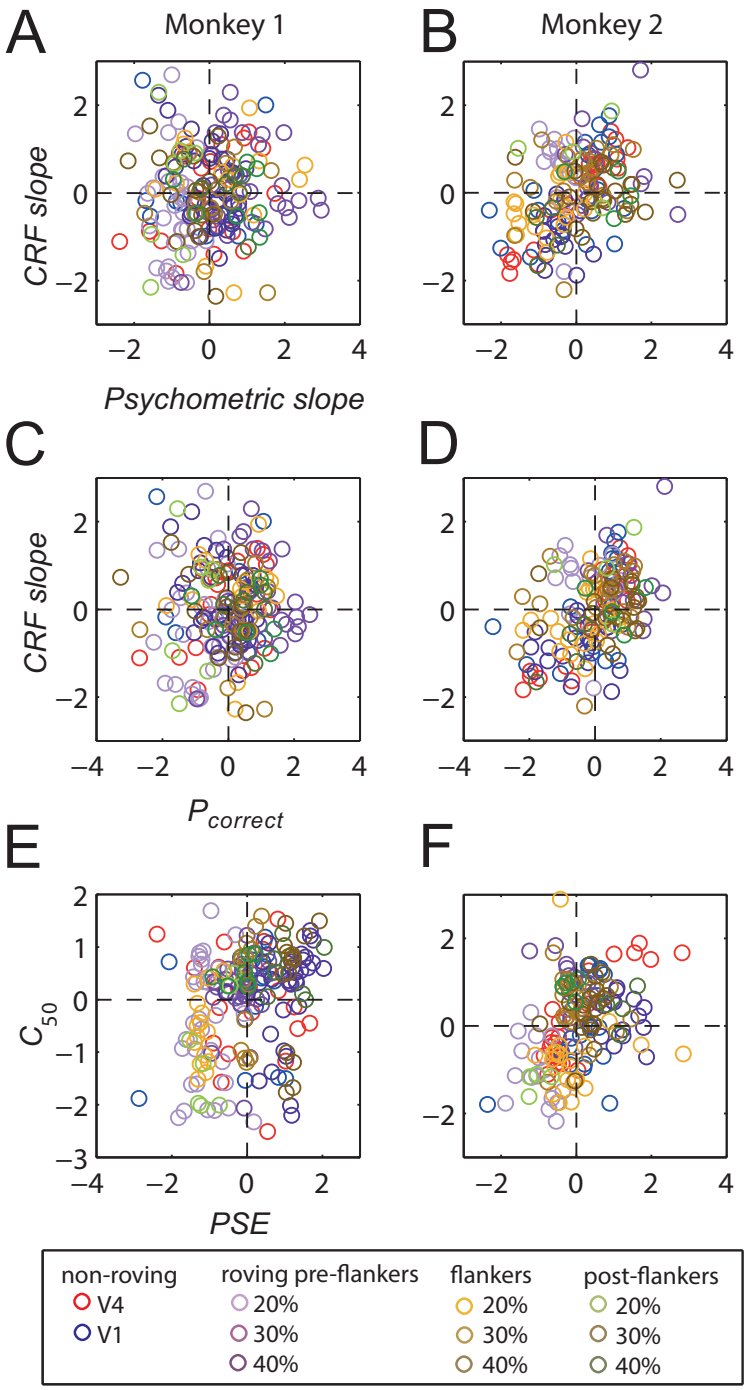


Figure 64. Plots of *z*-scored CRF parameters against *z*-scored psychometric function parameters for the entire training period, across V4 and V1 locations and across non-roving and roving sessions (colour coded by task paradigm). First column: monkey 1; second column: monkey 2. A & B: CRF slope against psychometric function slope; C & D: CRF slope against $P_{correct}$; E & F: C_{50} against the PSE.

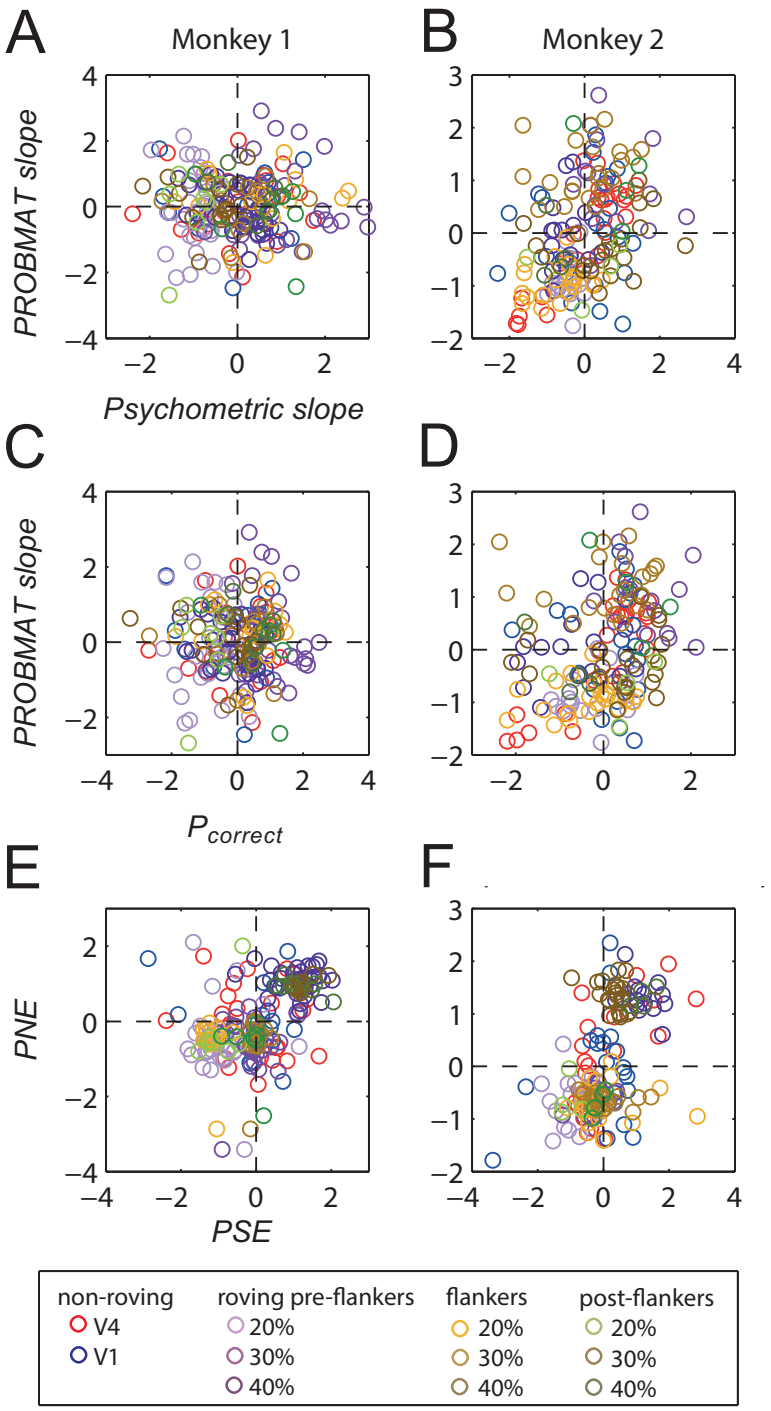


Figure 65. Plots of *z*-scored PROBMAT function parameters against *z*-scored psychometric function parameters for the entire training period, across V4 and V1 locations and across non-roving and roving sessions (colour coded by task paradigm). First column: monkey 1; second column: monkey 2. A & B: PROBMAT slope against psychometric function slope; C & D: PROBMAT slope against $P_{correct}$; E & F: PNE against the PSE.

A Spearman's rank correlation was carried out between each pair of neurometric and psychometric *z*-scored parameters of interest, to identify relationships between neuronal activity and the monkeys' response (Table 39).

	Monkey 1			Monkey 2		
Comparison	*df*	*r*	*q*	*df*	*r*	*q*
CRF slope vs psychometric slope	205	.0568	.423	171	.386	< .001*
CRF slope vs $P_{correct}$	205	-.040	.572	171	.418	< .001*
C_{50} vs PSE	205	.037	.601	171	0.386	< .001*
PROBMAT slope vs psychometric slope	205	-.045	.518	171	.362	< .001*
PROBMAT slope vs $P_{correct}$	205	.311	< .001*	171	.425	< .001*
PNE vs PSE	205	.478	< .001*	171	.485	< .001*

* $q < \alpha$

Table 39. Positive correlations between *z*-scored neurometric and psychometric function parameters were observed throughout non-roving and roving training (FDR correction for multiple comparisons, $\alpha = .05/12 \times 6 = .025$).

The slope of the PROBMAT function was positively correlated with $P_{correct}$ in both animals, and the slope of the PROBMAT function was positively correlated with the slope of the psychometric function in monkey 2. Thus, higher performance at the behavioural level was associated with greater discriminability in the neuronal responses to sample and test stimuli.

The PNE and PSE were also positively correlated in both monkeys. This revealed a systematic relationship between the contrast levels at which monkeys reported the test and sample stimuli as being identical, and the contrast levels at which sample- and test-evoked firing rates were indistinguishable, from the standpoint of an ideal observer.

The slope of the CRF at 30% was positively correlated with psychometric slope as well as with psychometric performance ($P_{correct}$) in monkey 2. This indicated that the better the animal's performance at the behavioural level, the steeper the CRF. The C_{50}

was also positively correlated with the PSE, in this subject. No significant correlation was seen between CRF parameters and psychometric parameters in monkey 1.

It was possible that modulatory effects on spiking activity differed between V4 and V1. If so, then the combination of data across both regions might mask any effects that occurred in opposite directions between the areas. Thus, this analysis was carried out separately for each of the recording locations. For the analysis involving only V1 sessions, *z*-scored data were pooled across non-roving and roving, flanker and flankerless paradigms. For the analysis involving only V4 sessions, data were confined to those acquired during non-roving training.

When V4 sessions were excluded from the analysis, results were qualitatively very similar to those obtained when both V4 and V1 sessions were included (Table 40). The only difference was the appearance of a positive correlation between the C_{50} and the PSE, in monkey 2.

	Monkey 1			Monkey 2		
Comparison	*df*	*r*	*q*	*df*	*r*	*q*
CRF slope vs psychometric slope	183	.009	.900	146	.283	< .001*
CRF slope vs $P_{correct}$	183	-.067	.364	146	.329	< .001*
C_{50} vs PSE	183	.004	.956	146	.304	< .001 *
PROBMAT slope vs psychometric slope	183	-.071	.335	146	.282	< .001*
PROBMAT slope vs $P_{correct}$	183	.374	< .001*	146	0.303	< .001*
PNE vs PSE	183	.565	< .001*	146	0.477	< .001*

* $q < \alpha$

Table 40. Positive correlations between *z*-scored neurometric and psychometric function parameters were observed throughout non-roving and roving training when stimuli were positioned at the V1 location, though this was true for more parameters in monkey 2 than in monkey 1 (FDR correction, $\alpha = .05/12 \times 7 = .0292$).

When only V4 data were included in the analysis, the results differed substantially for monkey 1: none of the neurometric parameters were significantly correlated to psychometric parameters (Table 41). Thus, it appeared that the positive

correlations seen between PROBMAT parameters and behavioural performance in monkey 1 stemmed primarily from the V1, rather than the V4, component.

In monkey 2, on the other hand, all of the neurometric and psychometric parameters were positively correlated (the correlation was significant for all the comparisons except for that between the PNE and the PSE), indicating that neuronal responses in V4 were closely linked to his behavioural performance.

	Monkey 1			Monkey 2		
Comparison	*df*	*r*	*q*	*df*	*r*	*q*
CRF slope vs psychometric slope	20	.392	.0718	23	.839	< .001*
CRF slope vs $P_{correct}$	20	.180	.421	23	.725	< .001*
C_{50} vs PSE	20	.335	.128	23	.790	< .001*
PROBMAT slope vs psychometric slope	20	.146	.514	23	.712	< .001*
PROBMAT slope vs $P_{correct}$	20	-.235	.290	23	.535	.00662*
PNE vs PSE	20	-.055	.809	23	.402	.0476

* $q < \alpha$

Table 41. Positive correlations between *z*-scored neurometric and psychometric function parameters were present throughout non-roving training for monkey 2, though not for monkey 1, when stimuli were positioned at the V4 location (FDR correction, $\alpha = .05/12\times5 = .0208$).

3.6.1.1 Discussion of correlations between neuronal activity and behaviour

In summary, numerous correlations were found between neurometric and psychometric measures in both monkeys, and the patterns observed were in the directions expected. Enhanced performance on the CD task was clearly associated with steeper neurometric and contrast response functions, while the location of the PSE (relative to the sample contrast) was predictive of the locations of the PNE in both monkeys, and of the C_{50} in monkey 2.

This overarching pattern was borne out in the observations made at each stage of training, whether in the absence or presence of flankers, or with non-roving or roving stimuli. An important (if complicating) feature of our findings was the difference in the

reactions of our two subjects to the insertion of flankers. In hindsight, this allowed us to study changes in neuronal activity under a variety of conditions- not only when subjects showed consistent improvement, but also when they were stymied by the task. Furthermore, although spiking activity underwent modulations in different directions between subjects at various stages of roving training, this could ultimately be explained by the fact that their neuronal activity depended on how much success they achieved on the task. When task performance was taken into account, the relationships between neurometric and psychometric measures of performance were consistent between subjects, regardless of whether advances or declines were made at any given stage of perceptual learning.

Chapter 4: Control tasks/ analyses

4.1 Roving task training with matching locations between the two monkeys

During training on the roving task, the RF locations of recorded V1 neurons differed slightly between the two subjects. The stimulus location used in monkey 1 (4.6° eccentricity) differed from that used in monkey 2 (1.5° eccentricity). This naturally raised the question of whether stimulus eccentricity contributed to the divergence in performance between subjects that was observed upon the introduction of flankers, and upon levels of roving performance in general.

To explore this possibility, an additional period of training was carried out. During these sessions, monkey 2 was presented with stimuli that were located at the same coordinates as those used for monkey 1, i.e. 4.6° from the centre of the visual field. Behavioural performance was monitored throughout, but no neuronal recordings were made as the stimuli used during this control task were no longer positioned within neuronal RFs.

4.1.1 Methods

4.1.1.1 Stimuli used in the control roving task

The stimuli used for monkey 2 at this stage of training were identical to those that were previously used for training at the V1 location in monkey 1 (see Table 42 for details).

Property	Monkey 2		
	Pre-flanker	Flankers	Post-flankers
No. of sessions	23	22	5
Location	parafoveal	parafoveal	parafoveal
Coordinates of centre (dva)	(-3.5, -3)	(-3.5, -3)	(-3.5, -3)
Size (dva)	3	3	3
SF (cpd)	2	2	2
Orientation	vertical	vertical	vertical
Stimulus type	sinusoidal grating	sinusoidal grating	sinusoidal grating
Flankers	absent	present	absent

Table 42. Stages of roving training and a list of stimulus properties, when stimuli were at the control location.

4.1.2 Results

This control task was carried out with monkey 2, for a total of 60 sessions. During the 'pre-flanker' sessions, training on a roving stimulus task was performed in the absence of flankers, for 23 sessions (over a period of 5 weeks). Next, the task was performed in the presence of flankers, for 22 sessions (4 weeks). Finally, during 'post-flanker' training, flankers were removed and training was carried out for 5 sessions (1 week).

4.1.2.1 Training at the control location during the pre-flanker period

Performance in terms of the proportion of correct trials, the PSE and the slope of the psychometric function were plotted against time (Figure 66). As before, during each stage of training, task performance, M, was compared between the first and last 30% of sessions (M_{early} and M_{late}) for each sample contrast. In the absence of flanker stimuli, the proportion of correct trials and the slope of the psychometric function increased significantly, and the PSE shifted towards the sample contrast, when the sample contrast was 20% and 40% (refer to Table 43). This pattern of task performance, in which

improvement was observed for selected measures and sample contrasts, resembled the pattern seen in either subject when roving stimuli were first introduced, prior to the control task.

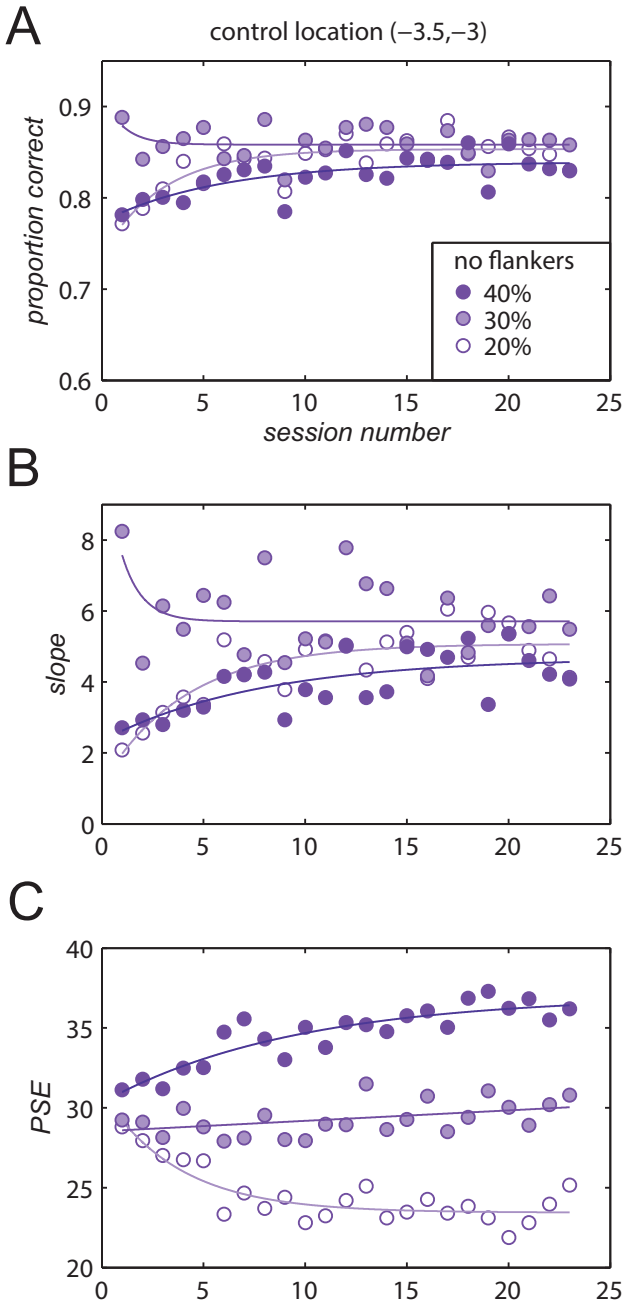


Figure 66. Performance during training with monkey 2 on the roving task, in the absence of flankers, when stimuli were placed at the control location. A: $P_{correct}$; B: slope; C: PSE.

Statistic	*df*	*t*	*q*
	20% sample		
$P_{correct}$	1,10	9.9	.0103*
Slope	1,10	15.3	.00291*
PSE	1,10	19.1	.0014*
$RT_{correct}$	1,10	0.4	.5346
RT_{error}	1,10	1	.3337
	30% sample		
$P_{correct}$	1,10	0.2	.6407
Slope	1,10	0.8	.395
PSE	1,10	2.8	.123
$RT_{correct}$	1,10	0.7	.4315
RT_{error}	1,10	1.3	.277
	40% sample		
$P_{correct}$	1,10	11.9	.0062*
Slope	1,10	14.5	.00345*
PSE	1,10	37.3	< .001*
$RT_{correct}$	1,10	1.4	.2615
RT_{error}	1,10	0.1	.7235

* $q < \alpha$

Table 43. Comparisons of performance between early and late sessions in monkey 2 during pre-flanker training, when stimuli were presented at the control location. The proportion of correct trials ($P_{correct}$) and the slope of the psychometric function increased significantly with training, and the PSE shifted towards the sample contrast values, for the 20% and 40% sample conditions (Student's *t*-test, FDR correction for α-levels, proportion correct: α =.05 × 2/3 = .0333; slope: α =.05 × 2/3 = .0333; PSE: α = .05 × 2/3 = .0333; $RT_{correct}$: α =.05/3 = .0167; RT_{error}: α = .05/3 = .0167).

4.1.2.2 Addition of flanker stimuli at the control location

Over the course of training with flanker stimuli, a significant improvement occurred for all three sample contrasts in terms of the proportion of correct trials and the slope of the psychometric function, and a shift in the PSE occurred towards the value of 30%, for the 30% sample contrast condition (Figure 67 and Table 44).

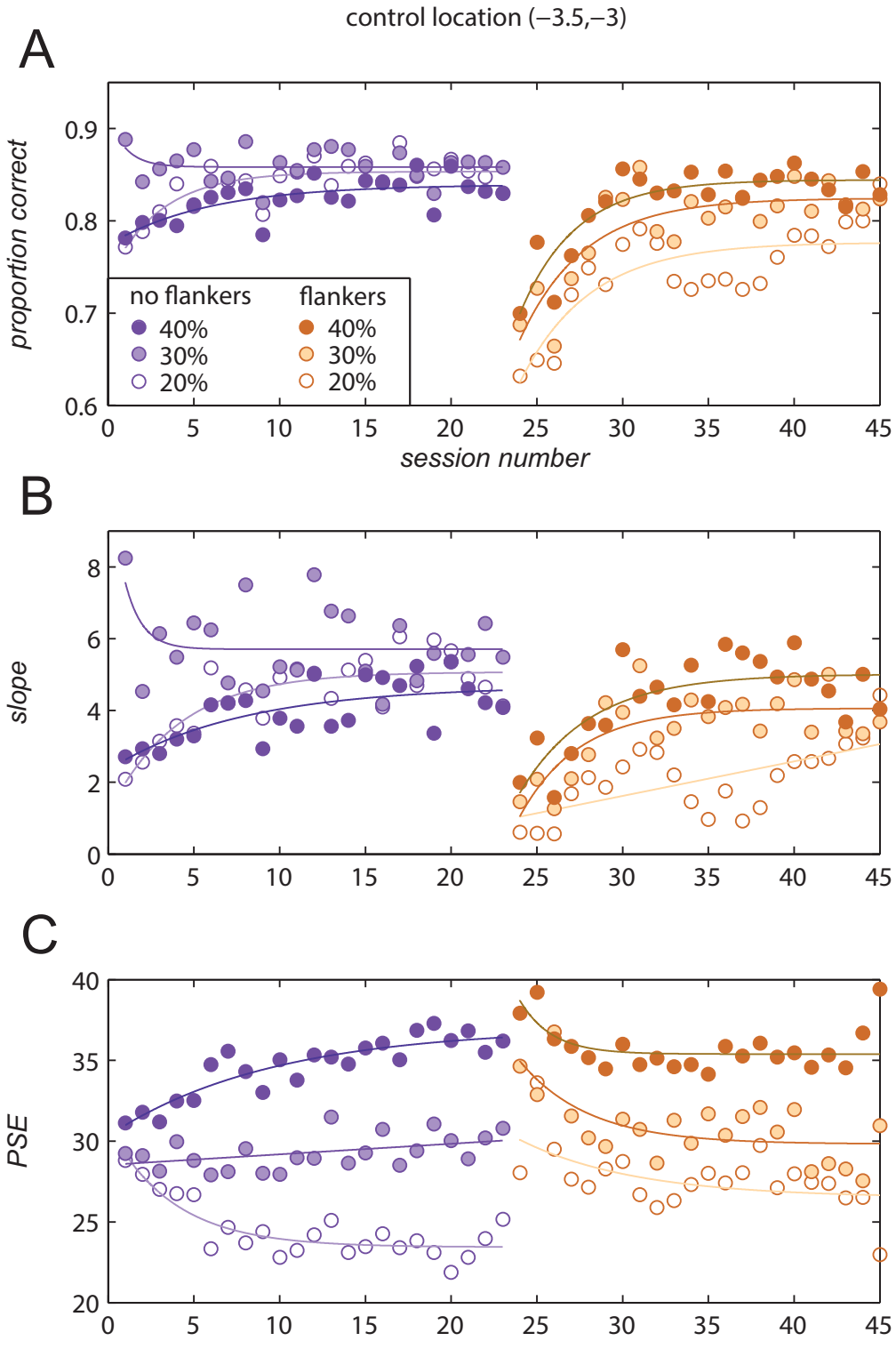


Figure 67. Performance during training with monkey 2 on the roving task, in the presence of flanker stimuli, at the control location (orange markers). Previous levels of performance (in the absence of flankers) are also depicted for comparison (purple). A: $P_{correct}$; B: slope; C: PSE.

Statistic	*df*	*t*	*q*
	20% sample		
$P_{correct}$	1,10	22.5	< .001*
Slope	1,10	20	.0012*
PSE	1,10	4.5	.0603
$RT_{correct}$	1,10	5.4	.0423
RT_{error}	1,10	7.4	.0212
	30% sample		
$P_{correct}$	1,10	14.2	.0036*
Slope	1,10	9.2	.0125*
PSE	1,10	6.5	0.029
$RT_{correct}$	1,10	8.2	.0167*
RT_{error}	1,10	3.2	.103
	40% sample		
$P_{correct}$	1,10	13.2	.0046*
Slope	1,10	15.6	.00274*
PSE	1,10	0.2	.651
$RT_{correct}$	1,10	9.2	.0126
RT_{error}	1,10	5.6	.0399

* $q < \alpha$

Table 44. Comparisons of performance between early and late sessions in monkey 2, during flanker training, when stimuli were presented at the control location. $P_{correct}$ and the slope improved across all three sample contrast conditions. Improvements in the PSE and RT were also seen on for some sample contrasts (Student's *t*-test, FDR correction for α-levels, proportion correct: α =.05 × 3/3 = .05; slope: α =.05 × 3/3 = .05; PSE: α =.05/3 = .0167; $RT_{correct}$: α =.05/3 = .0167; RT_{error}: α = .05/3 = .0167).

The control task was carried out to determine whether the differences in performance between two monkeys (seen during the first instance of roving training, prior to the control task) were induced by differences in stimulus parameters and in stimulus location. A comparison of performance levels between pre-flanker and flanker sessions revealed that although performance had increased within the period of flanker training itself, the addition of flankers had caused a significant drop in performance. Any improvements seen during flanker training contributed to a recovery in performance to pre-flanker levels, and the subject never managed to improve beyond the maximum levels attained during pre-flanker training (Table 45). The pattern seen at this location in monkey 2 thus matched the pattern seen during roving training prior to the control task, and differed from that observed in monkey 1.

Late pre-flanker performance versus late flanker performance			
Statistic	*df*	*t*	*q*
	20% sample		
$P_{correct}$	1,10	31.2	< .001*
Slope	1,10	32.0	< .001*
PSE	1,10	17.0	.00205*
$RT_{correct}$	1,10	4.2	0.0675
RT_{error}	1,10	6.0	.0347*
	30% sample		
$P_{correct}$	1,10	11.7	.00650*
Slope	1,10	18.7	.00149*
PSE	1,10	0.3	0.598
$RT_{correct}$	1,10	7.0	.0246*
RT_{error}	1,10	3.5	0.0908
	40% sample		
$P_{correct}$	1,10	0.0	0.934
Slope	1,10	0.0	0.837
PSE	1,10	0.1	0.741
$RT_{correct}$	1,10	7.7	.0198*
RT_{error}	1,10	2.4	0.1533

* $q < \alpha$

Table 45. Comparisons of performance of monkey 2 between late pre-flanker and late flanker sessions (Student's *t*-test, FDR correction for α-levels, proportion correct: α =.05 × 2/3 = .0333; slope: α =.05 × 2/3 = .0333; PSE: α =.05 × 1/3 = .0167; $RT_{correct}$: α =.05 × 2/3 = .0333; RT_{error}: α = .05 × 1/3 = .0167).

4.1.2.3 Removal of flankers from the control location

This stage of post-flanker training consisted of five sessions in which monkey 2 practised a flankerless CD task with roving stimuli positioned at the control location, immediately after undergoing training in the presence of flankers (Figure 68). The performance observed in the absence of flankers during the last of the control post-flanker sessions was compared to that attained during late control pre-flanker sessions. As with the comparison of performance made between pre-flanker and flanker sessions conducted prior to the control task, this step allowed us to determine whether flanker training had any enduring effects on performance of the CD task.

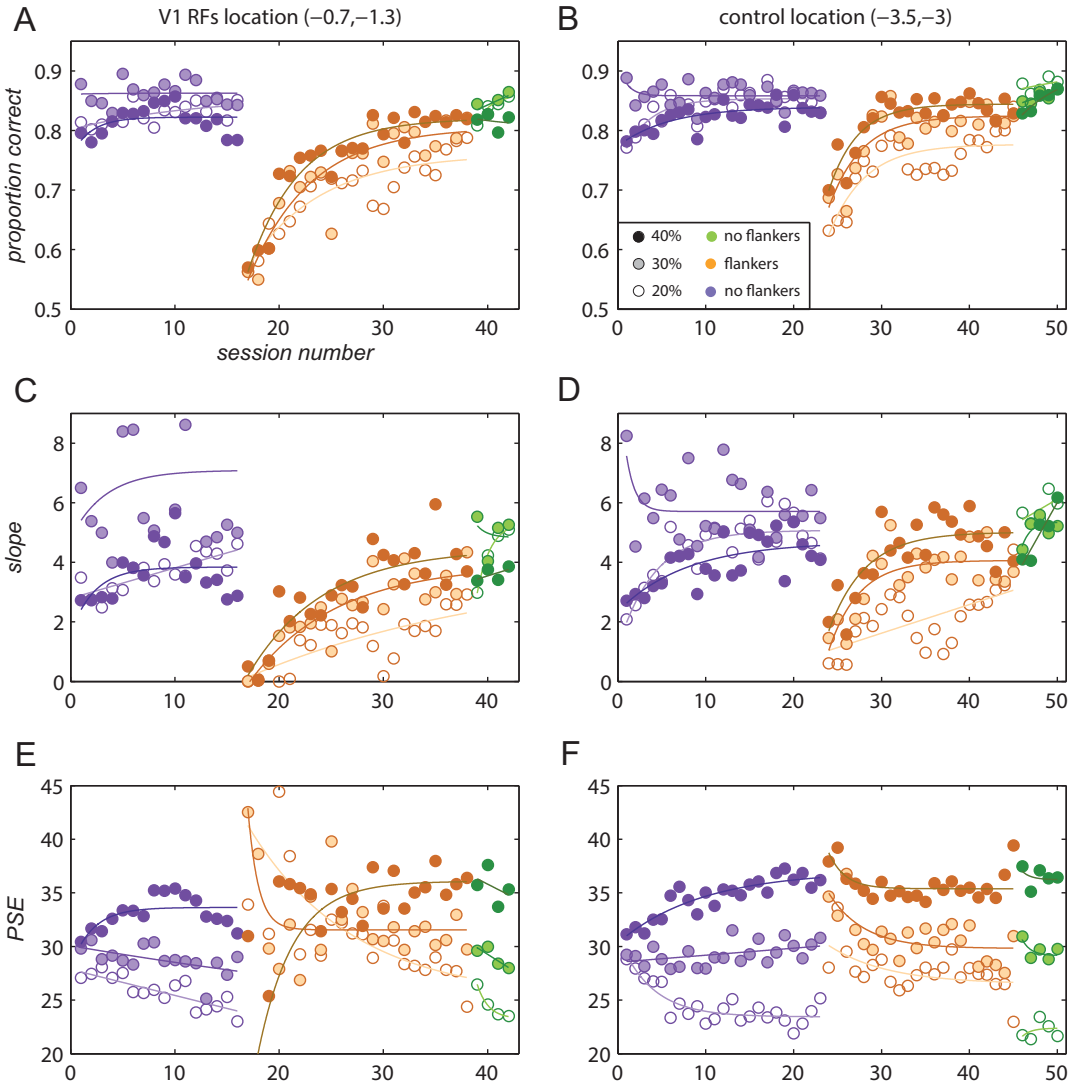


Figure 68. Overall performance of monkey 2 during his two versions of the roving task. Left column: performance on the roving task when stimuli were located just outside the fovea (the data are reproduced from Figure 61, page 174); right column: performance at the control location. A & B: $P_{correct}$; C & D: slope of the psychometric function; E & F: PSE. Purple data points: pre-flankers; orange data points: flankers; green data points: post-flankers. Unfilled markers: 20% sample contrast conditions; medium-coloured filled markers: 30%; dark-coloured filled markers: 40%.

For the most part, monkey 2's performance during this control session (X_c) fell within the ranges of values seen during the late phase of pre-flanker training (Table 46). The proportion of correct responses, the slope, the PSE, $RT_{correct}$ and RT_{error} lay within the ranges attained during the late phase of pre-flanker training for the 20% sample, while they were either within the range or lay close to it, for the 30% and 40% sample.

Thus, the subject's ability to make fine contrast discriminations in the absence of flankers was largely comparable between pre- and post-flanker training, indicating that the dips in performance seen during the addition of flankers were temporary and only occurred in the presence of flankers.

Performance	Late pre-flanker sessions, range $X_{min} - X_{max}$	Last post-flanker session, X_c
	20% sample	
$P_{correct}$ (%)	75.2 – 82.5	76.3
Slope	3.5 – 6.6	6.0
PSE	27.7 –34.8	28.7
$RT_{correct}$	100.1 –124.3	119.4
RT_{error}	110.6 –148.5	136.9
	30% sample	
$P_{correct}$ (%)	78.9 –88.6	77.7
Slope	4.8 –6.4	5.2
PSE	28.7 –34.7	32.4
$RT_{correct}$	103.0 –118.9	120
RT_{error}	113.5 –143.7	131.6
	40% sample	
$P_{correct}$ (%)	79.5 –83.3	80.5
Slope	3.7 – 6.0	6.2
PSE	33.1 –37.4	34.8
$RT_{correct}$	102.6 –121.9	123.4
RT_{error}	91.7 –136.9	115.3

Table 46. A comparison of monkey 2's performance on the control task, during post-flanker sessions, and during the end of pre-flanker training. $X_{min} - X_{max}$: Ranges of performance seen during late pre-flanker sessions. X_c: Performance recorded during the last session of the post-flanker task.

4.1.3 Summary of results from the roving task at the control location

Uneven gains in performance during pre-flanker training were followed by an initial drop in performance when flankers were introduced, but this was then followed by consistent improvements across all three sample contrasts during flanker training. This pattern of performance on the control task was essentially identical to that seen when stimuli were located closer to the fovea, prior to the control task, when monkey 2 transitioned from a pre-flanker to a flanker paradigm.

Thus, the differences in performance seen between the two subjects during pre-flanker and flanker stages of training were not simply due to differences in stimulus eccentricity. In monkey 1, the initial drop in performance that occurred during the first session with flanker stimuli was followed by an immediate recovery and rapid improvement over the course of training with flankers (refer to Figure 60). Upon removal of flankers, monkey 1's performance returned to lower, pre-flanker levels. In monkey 2, on the other hand, performance levels dropped sharply upon the introduction of flankers and failed to surpass those seen during flankerless training, regardless of stimulus location. Upon removal of flankers, performance levels returned to or even slightly exceeded those of pre-flanker training (refer to Figure 68).

4.1.4 Possible differences in task strategy

Thus, it appears that the two subjects differed systematically in their reaction to flanker stimuli, and possibly in their approach to the task. It is conceivable that while monkey 1 maintained a focus on making a comparison between sample and test stimuli, and used the flanker stimuli as aids in making the comparison, monkey 2 perceived the flankers as being part of the stimuli that were to be compared- in other words, the addition of flankers effectively increased the noise levels in the visual signal (refer to Figure 69 for an illustration of this hypothetical scenario).

The top row of Figure 69 depicts sample and test stimuli (C_s = 20% contrast and C_t = 5% contrast, in this example) that are presented in the absence of flankers. During the presentation of the sample, the subject has to observe and note its contrast. Once the sample stimulus disappears, the subject has to rely on a memory trace, in order to make a comparison between the memory of the sample contrast and the contrast of the test (this is termed the 'memory strategy'). This strategy requires a comparison to be made across stimuli that are separated by a gap in time. The integrity of the memory trace suffers from the processes of degradation (when the stored image fades from memory and gets distorted by noise) and adaptation (when the visual system adapts to the presentation of the sample and fails to respond as strongly to the test).

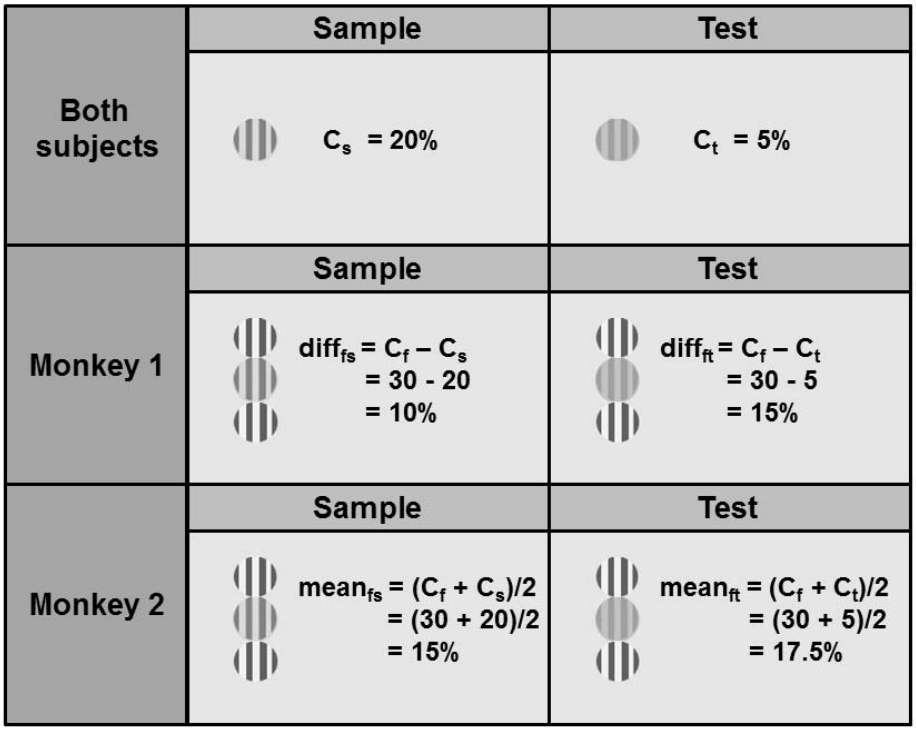


Figure 69. Illustration of possible strategies that might have been used by the subjects to carry out the contrast discrimination task.

The middle row depicts the hypothetical situation in which the subject is able to rely not only on the 'memory strategy' described above, but additionally, on a strategy that involves the comparison of two sets of simultaneously-presented stimuli. When the sample appears, accompanied by flankers, the subject compares the contrast of the sample with that of the flankers (in the example depicted, the difference in contrast, $diff_{fs}$ = +10%). When the test appears, also accompanied by flankers, which are identical in contrast to those that accompanied the sample, the subject compares the contrast of the test with that of the flankers ($diff_{ft}$ = +15%). He is now able to supplement the retrieved information about the contrast of the sample with additional information about the differences in contrast between the central and flanking stimuli. By comparing the contrasts of the sample and flanker stimuli, he calculates that the sample is 10% lower in contrast. When he subsequently compares the contrasts of the test and flanker stimuli, he realises that the difference between this set of stimuli (15%) is greater than the difference between the previous set, thus the test must be of lower contrast than the sample (the 'difference strategy'). Each judgment is made based on

simultaneously-presented stimuli and although the memory trace of $diff_{fs}$ from the first set is subject to degradation and adaptation, as with C_s in both the flanker-absent and the flanker-present task paradigms, the addition of useful information is likely to help, rather than hurt, his decision.

The bottom row depicts a hypothetical (and counter-productive) strategy that leads to poorer task performance. If, instead of distinguishing between the central and flanking stimuli, the subject proceeded to merge the stimuli into one perceived stimulus, then the contrast differences between central and flanker stimuli are not exploited in the way they would be with the difference strategy; rather, the contrasts of all three simultaneously-present stimuli are averaged out (the 'mean strategy'). As the flanker gratings are always presented at a contrast level of 30%, any averaging that includes the flanker contrasts will always result in a reduction of the difference between the first and second sets of 'aggregated' stimuli.

While it is not possible to determine unequivocally whether the disparity in performance between the subjects was due to a difference in task strategy, or to verify whether the task strategies outlined in Figure 69 accurately describe those adopted by each subject, this theory provides a plausible explanation for the marked differences observed between the subjects.

4.1.4.1 Possible effects of the order of exposure to training stimuli

Monkey 2 was exposed to relatively tiny grating stimuli at an eccentricity of 1.5°, before being exposed to the stimuli of his control task, which were larger and located at 4.6° eccentricity. It is possible that the sequence of exposure he received was a contributing factor to his performance. If he had learnt to chunk the parafoveally located stimuli into one perceived stimulus, and hence adopted this strategy when performing the task at the control location, he may not have had the chance to develop the 'difference approach' and form judgments based on discriminations between simultaneously-presented stimuli. One wonders whether monkey 1 might have displayed a similar learning pattern (or lack thereof) to that of monkey 2, had he been trained on a task with stimuli located at a lower eccentricity, before undergoing training on the task with stimuli at an intermediate eccentricity.

4.2 Spatial attention control task

Attention exerts modulatory effects on the CRF, which are represented by the response gain, contrast gain, and additive models of attention (Buracas & Boynton, 2007; Thiele et al., 2009; Williford, 2006). One could argue that the shifts in PNE that were observed in V4 over the course of learning might not have been due specifically to improvements at the perceptual level on the CD task, but rather to a general effect of attention. If, for example, top-down attention triggered a shift in the PNE towards the stimulus contrast, and this effect was strengthened as a result of training, then one might see such a shift due to the tuning of mechanisms at higher levels of the cognitive hierarchy, without the direct involvement of areas such as V1 and V4. In this scenario, the focusing of attention upon the stimuli presented in the CD task would be enough to trigger a shift in the PNE, as long as subjects had gained sufficient familiarity with the task.

To address this issue, a control task was performed with monkey 2 to investigate whether the presence of spatial attention affected contrast-dependent responses to the stimuli used during training; specifically, we wanted to determine whether it was able to induce a shift in the location of the PNE of the PROBMAT function.

4.2.1 Methods

During this stage, two sets of visual stimuli were shown onscreen simultaneously- one set was located in the lower left visual field (within either the V4 or V1 neuronal RFs), while the other was located in the upper visual field (outside the RFs). Stimuli in the RFs always consisted of vertically-oriented sinusoidal gratings of varying contrast, whereas stimuli outside the RFs always consisted of pairs of sinusoidal gratings at 96% contrast (one vertically oriented, the other horizontally oriented).

For one half of each recording session, the animal had to attend to stimuli within the RFs and perform a CD task, in which he discriminated between two sequentially-presented stimuli of different contrasts. During the other half of the session, he had to attend to stimuli outside the RFs and perform an orientation discrimination task (Figure 70).

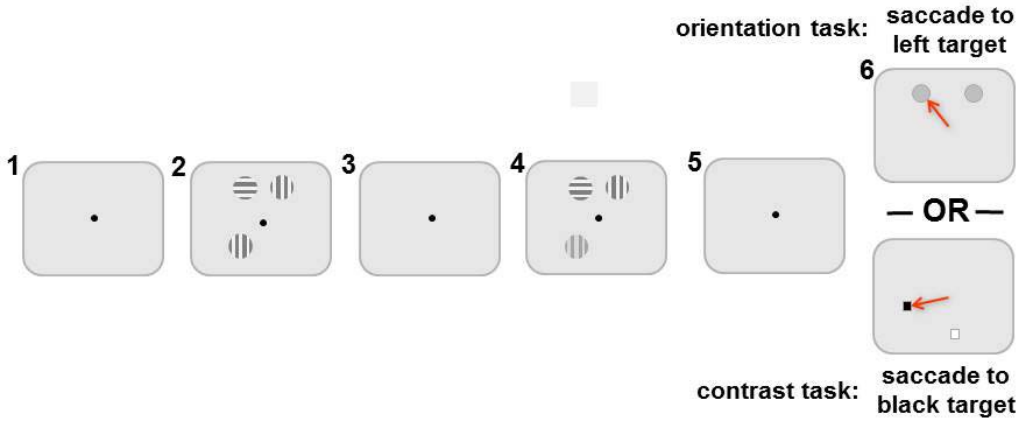


Figure 70. Control task performed by monkey 2, to direct spatial attention at or away from neuronal RFs. During one half of each recording session, the subject had to attend to a pair of gratings in the upper visual field, and report the location of the horizontal grating. During the other half of the session, he had to attend to stimuli that appeared in the lower visual field in order to perform a contrast discrimination task.

4.2.2 Results

As with the previous sets of ROC analysis, cumulative PROBMAT values were calculated across channels, based on levels of spiking activity that were elicited by the sample and test stimuli, yielding a measure of how well neurons discriminated between various stimulus contrasts (Figure 71). This was done separately for stimuli presented during the CD task (when attention was directed to the RFs) and during the orientation discrimination task (when attention was diverted away from the RFs).

A three-way ANOVA was performed, with the locus of spatial attention (within or outside the RFs), session number, and condition number as factors. A significant main effect of attention was observed at both recording sites (V1: $F(1,94) = 91.3$, $p < .001$; V4: $F(1,202) = 8.5$, $p = .0039$). Post-hoc tests revealed that for the V1 location, PROBMAT values were substantially higher when attention was within the RFs and the monkey performed a CD task, than when attention was outside the RFs and the monkey performed an orientation discrimination task. This corresponded to a leftward shift of the PROBMAT curve. At the V4 location, on the other hand, PROBMAT values were somewhat smaller for low contrasts, when attention was within the RFs (corresponding to a downward shift in the range of the PROBMAT function).

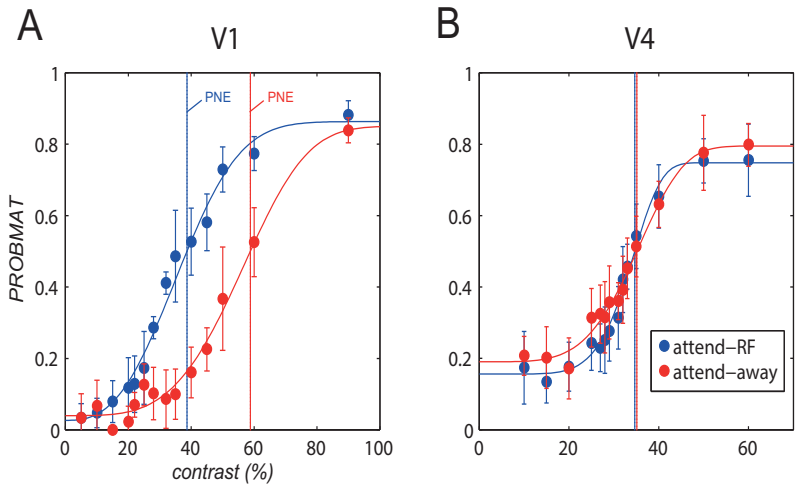


Figure 71. Distributions of PROBMAT values during attend-RF (blue) and attend-away (red) perceptual tasks in V1 (A) and V4 (B). Error bars show the SD across sessions (V1: $N = 4$; V4: $N = 8$). Vertical lines indicate the PSE (V1, attend-RF: 38.6%, attend-away: 57.6%; V4: attend-RF: 35.5%, attend-away: 35.3%).

The presence of spatial and task-dependent attention was thus visible as a shift of the PNE in the V1 response, whereas it had little effect on the PNE in V4. As this control experiment was carried out after training, it is not possible to determine to what extent modulations observed in V1 were caused by learning, and to what extent they were due to effects of attention. However, at the V4 location, significant shifts in the population PNE had occurred over the course of training (refer to the section titled, '*Changes of the population neurometric function with training*' on page 79), and based on the results from the control task, shifts could not be attributed solely to the engagement of attention. Thus, it was likely that the change in the PNE that was seen in V4 was indeed induced by perceptual learning, and was not merely an attention-induced artifact.

Final discussion and further work

The issue of whether substantial improvement in contrast discrimination is possible during adulthood, and the nature of circumstances that permit this, has been addressed in a number of human psychophysics studies (Adini et al., 2002; Dorais & Sagi, 1997; Kuai et al., 2005; Phan & Ni, 2011; Polat et al., 2004; Yu et al., 2004; J.-Y. Zhang et al., 2008). In tasks involving perceptual domains other than stimulus contrast, neuronal recordings in NHPs have revealed learning-induced changes in a variety of visual cortical areas, from low level (Crist et al., 2001; W. Li et al., 2004; Schoups et al., 2001) to intermediate and higher level regions (Baker et al., 2002; Law & Gold, 2008; Raiguel, 2006; Rainer et al., 2004; Yang & Maunsell, 2004; Zivari Adab & Vogels, 2011).

We observed improvements in psychometric performance as our adult macaque subjects underwent training on a contrast discrimination task, under both non-roving and roving conditions. Simultaneously, we recorded changes in activity from a stable subpopulation of neurons in striate and extrastriate cortex, which demonstrated that both V1 and V4 contribute to and are reflective of the animals' perceptual abilities at the behavioural level. The exact nature of effects seen at the neuronal level was closely coupled to the animals' behavioural performance. Table 47 presents a summary of the changes observed during training on the non-roving PL task, while Table 48 describes performance-dependent modulations of V1 activity, during training on the roving task. (A summary of the control tasks used during this experiment is provided in Table 49.)

In brief, correlations between the CRF, the PROBMAT function, and the psychometric function were visible across the training period, and the addition of flankers to the task paradigm exerted strong modulatory effects (albeit in different directions between the two monkeys) on task performance, which amplified the relationships between psychophysical and neuronal metrics. Across subjects, superior task performance was accompanied by wider ranges in the CRF and PROBMAT function, by steeper slopes of the CRF at the sample contrast, and by shifts in the C_{50} and the PNE towards the sample contrast. These corresponded to wider ranges in stimulus-evoked spiking activity and stimulus discriminability, and finer discriminative abilities at the contrast levels that were behaviourally relevant.

Perceptual learning task							
Behavioural		Neuronal					
Psychometric function		CRF			Neurometric function		
V4 (non-roving)							
$P_{correct}$	↑	Slope	Individual channels	↕ *M1;* ↑ *M2*	Slope	Individual channels	↕ *M1;* ↑ *M2*
Slope	↑		Population	↑ *M2*		Population	↑ *M2*
Threshold	↓ *M2; trend* ↓ *M1*	-			Threshold	Population	↓ *M2*
PSE	→ *30% M2; already close to 30% in M1*	C_{50}	Individual channels	*Mostly* → *30%*	PNE	Individual channels	→ *30%*
			Population	→ *30% M2*		Population	→ *30%*
RTs	↓	-					
V1 (non-roving)							
$P_{correct}$	↑	Slope	Individual channels	*Mostly* ↑ *M2*	Slope	Individual channels	↑ *M2*
Slope	↑		Population	↑ *M2*		Population	*No change*
Threshold	↓	-			Threshold	Population	↓ *M2*
PSE	*No change*	C_{50}	Individual channels	↔ *30% M2*	PNE	Individual channels	→ *30%*
			Population	← *30% M2*		Population	*Trend* → *30% M1*
RTs	↓	-					

Table 47. Summary of behavioural and neuronal changes during PL on the non-roving task in V4 and V1. ↑: increase occurred; ↓: decrease occurred; ↕ both increases and decreases occurred, depending on the channel; → *30%*: shift occurred towards 30%; ← *30%*: shift occurred away from 30%; ↔ *30%*: shifts occurred both towards and away from 30%, depending on the channel. M1: monkey 1; M2: monkey 2. '*Trend*' indicates that a shift was observed, but was not significant.

Performance-linked neuronal properties in V1							
Good performance				Poor performance			
CRF		Neurometric function		CRF		Neurometric function	
Slope	↑	Slope	↑	Slope	↓	Slope	↓
C_{50}	→ *sample contrast*	PNE	→ *sample contrast*	C_{50}	← *sample contrast*	PNE	← *sample contrast*
Range	↑	Range	↑	Range	↓	Range	↓
Fano factor				Fano factor			
↑				↓			

Table 48. Summary of the performance-dependent characteristics of the CRF, the neurometric function, and the Fano factor, observed across the population of V1 neurons, during the roving task. (Note that the emergence of these modulations were not necessarily linked to PL of the CD task, but could have been triggered by a combination of factors such as attention modulation and subject-specific task strategy). ↑: higher; ↓: lower; → *sample contrast*: value lay closer to the sample contrast; ← *sample contrast*: value lay further away from the sample contrast.

Control tasks				
Location	Original	Manipulation	Outcome	Section
V4	Vertically oriented Gabors	Horizontally oriented Gabors	No effect, i.e. transfer of learning to a novel stimulus orientation occurred	Control task with horizontally-oriented Gabor stimuli at the V4 location, page 35
V4	Gabor stimuli	Grating stimuli	No effect, i.e. subjects relied on CD rather than perceived size	Control task with sinusoidal grating stimuli at the V4 location, page 36
V1	SF 4	SF 2	Worsening in performance, i.e. transfer of learning to novel SF was limited	Control task with stimuli of different spatial frequencies at the V1 location, page 37
V1	Sample present	Sample absent	No change in PSE, i.e. internalised 30% contrast was used as a reference	Control task with only the test stimulus- not the sample- at the V1 location, page 37
V1	Flankers present	Flankers removed	Performance returned to pre-flankers levels, i.e. changes during flanker training only occurred in the presence of flankers	Removal of flanker stimuli, page 172
V1	1.5° eccentricity	4.6° eccentricity	No change in pattern of results in monkey 2, i.e. differences at the neuronal level between the two monkeys were not solely attributable to differing stimulus eccentricities	Roving task training with matching locations between the two monkeys, page 190
V4	Spatial attention lay within RFs	Spatial attention lay outside RFs	No attention-induced shifts in PNE	Spatial attention control task, page 202
V1	Spatial attention lay within RFs	Spatial attention lay outside RFs	Increase in PROBMAT values and attention-induced leftward shift in PNE	

Table 49. Summary of the key control tasks used in this study, and their results and implications.

While behavioural improvements and their accompanying effects on neurometric performance were visible when performance was good, the same held true when performance was poor. Deteriorations in performance were accompanied by the reverse effects on V1 activity, including a decrease in the slope of the CRF, a shift in the C_{50} and PNE away from the sample contrasts, and a narrowing in the range of spiking activity.

During the non-roving task, these changes were closely coupled to the gains in performance that were seen in both subjects, demonstrating that the neural correlates of PL were present in V4; changes in V1, on the other hand, while clearly present, did not appear to be able to satisfactorily account for the behavioural improvements observed (a thorough discussion of the relative contributions of the two cortical regions is presented in the section titled, 'Discussion of neuronal results from the CD task,' page 109). The lack of transfer of learning to stimuli of an unfamiliar SF further supported the idea that V1 is less directly involved than V4. Task training was further characterised by changes in response adaptation (the directions of which may have signalled the adoption of differences in task strategies by the two subjects) and by increases in choice probability in V4 for both monkeys, and in V1 for monkey 2.

A variety of intriguing questions arise from this work: how do the different areas interact with each other to influence activity? What role does top-down attention play in the selection of sites of plasticity? What factors determine whether learning is enabled or disabled (e.g. duration of training, sequence of training paradigms, scope for chunking of sequences into memory, prior exposure to stimuli), and what are the neural mechanisms behind them? How do our findings relate to the broader context of perceptual learning; specifically, how do they support or refute the theories described in the Introduction ('Models of perceptual learning,' page 5)?

Our results support a model of PL of contrast discrimination in which neuronal plasticity occurs at both intermediate- and low-level regions, but it is the changes in V4 (at a minimum) that are most strongly correlated with improvements in performance. Thus, the 'early learning' model described in the introduction to Chapter 1 ('Early learning model,' page 6), in which changes are restricted to low-level cortical areas, provides an inadequate description of our findings in the domain of contrast

discrimination. The 'late learning' theory (page 8) emphasises learning-induced changes at intermediate- to high-level cortical regions, thus offering an improvement over the early learning model, but it still fails to capture or account for the 'collateral' changes in V1 that occur during training. The reverse hierarchy theory of learning (page 11) encompasses changes over a variety of levels in the visual hierarchy and provides the best description of our findings thus far; however, the predictions made by the model regarding the sequence of events across the hierarchy are too advanced to be tested in the current study. Our subjects undertook training with stimuli positioned first at the V4 location, then at the V1 location (potential effects of which are discussed in the section, 'Possible effects of the order of exposure to training stimuli,' page 201). Note that as there was no spatial overlap between our V4 and V1 RFs, we were unable to unequivocally rule out the possibility that PL-related changes occur in V1 during the early days of training; conservatively speaking, while our data suggest that V1 is not primarily responsible for the learning of fine discriminations, this argument is still open to debate. We could, in principle, posit a 'forward hierarchy theory of learning,' in which changes occur at low-level regions and progress through intermediate and then high-level ones; or a 'mixed hierarchy theory' in which changes occur simultaneously across multiple areas. Simultaneous recording of activity from these two areas during presentation of stimuli in overlapping RFs would allow a closer examination of the timeline of events across both V4 and V1.

Numerous studies have examined the effects of top-down attention on neuronal responses within short time intervals (i.e. across trials within a session). Attention effects on the CRF have been described by response gain, activity gain, and contrast gain models. We observed changes in the shape of the CRF in V4 with learning, through an increase in the range of spiking responses that corresponded to 10% to 60% contrast levels in V4. We also noted a decrease in the response range corresponding to 5% to 90% in V1 with training. As we did not measure the maximum size of responses to optimal stimuli, we were unable to verify to what extent these effects supported the predictions made by the respective models.

Furthermore, in our task, spike thresholds were deliberately selected such that levels of spontaneous activity were uniformly matched across sessions (refer to the section, '*Automated threshold setting to obtain uniform spontaneous activity levels*

across sessions,' page 218). The issue of threshold setting is one that affects every electrophysiology study in which neuronal activity is compared across multiple sessions; manual selection of thresholds from one recording day to the next is subject to human error and care must be taken to minimise the introduction of systematic biases. The benefit of our approach was that while we could not detect changes in levels of spontaneous activity with time, we had a fixed standard that allowed the robust identification of changes in stimulus-evoked activity, relative to spontaneous firing rates. However, this meant that we were unable to verify whether the changes observed in the range of the CRF were due to a response gain or an activity gain, as spontaneous activity levels may have changed during training without our knowledge.

Nevertheless, with respect to the steepening in the slope of the CRF, the modulatory effects of attention reported by earlier studies were qualitatively similar to the learning-induced changes in V4 which were observed in our task (aside from the fact that ours occurred over a much longer period). We found that improvements in performance were associated with a steepening of the CRF, a shift in the C_{50} of individual channels towards the sample contrast in V4 and V1, and increases in the variability of spiking responses. In the absence of large gains in performance, prolonged training was accompanied by gradual decreases in the maxima (relative to spontaneous levels) and the slope of the CRF, as well as by rightward shifts of the C_{50} away from the sample contrast in individual V1 channels. Perhaps during the early stages of learning, the changes that are initially enforced by attention become permanently encoded as a result of training, and the brain optimises and rewires its connections such that the task-specific 'spotlight' of attention no longer needs a switch, but is left permanently on. With extensive practice on the task (a form of over-training), the site of learning may shift elsewhere (such as towards the fine-tuning of connections between task-relevant neurons), contributing to the gradual declines in V1 individual channel activity observed in this study and in V1 population activity as reported by Ghose et al. (2002).

This leads to the question of whether conscious deployment of attention is required for this hypothetical process to occur. A number of studies have shown that under certain circumstances, passive viewing of behaviourally irrelevant stimuli is enough to boost subsequent performance on a perceptual task; in some cases, mere exposure to stimuli during engagement in an unrelated task facilitates transfer of

perceptual learning- a process known as 'task-irrelevant PL' (Seitz & Watanabe, 2003; Watanabe, 2001), while in others, training on a primary task along with training on an unrelated task can boost performance of the primary task at a secondary location (Xiao et al., 2008) or with untrained stimuli (J.-Y. Zhang et al., 2010). Indeed, Tartaglia, Bamert, Mast, and Herzog (2009) found that the reverse holds true- that in the absence of external stimuli, when subjects are instructed to *imagine* the appearance of stimuli using internal imagery, the preparatory stage of visualisation boosts subsequent performance on a perceptual task. A comparison of neuronal activity during initial exposure (or mental visualisation) to that during transfer of learning would shed light on the underlying cortical mechanisms.

Appendix A: Artifact removal from neuronal data

A.1 Generation of continuously-sampled channel data

Overall data processing (past the acquisition stage) was rather involved. This was due to various factors, not least to our desire to achieve comparable activity levels from each channel across recording days. Despite the complexity, the rationale behind all the different steps should become evident within the context of the following sections.

Raw data were acquired at a sampling frequency of 32556 Hz with a 24-bit analog-to-digital converter, with minimum and maximum input ranges of 11 and 136986 microvolts respectively (pre-set by Neuralynx, Inc.), a DMA buffer count of 128, and a DMA buffer size of 10 ms, using a 64-channel Digital Lynx 16SX Data Acquisition System (Neuralynx, Inc.). Digital referencing of voltage signals was performed prior to the recording of raw data, using commercially-provided Cheetah 5 Data Acquisition Software v. 5.4.0 (Neuralynx, Inc.), to yield good signal-to-noise ratios for each channel.

Following each recording session, the raw data were processed offline in a series of steps, using both commercial (Neuralynx, Inc.) and custom-written software.

In the first stage of processing, signals corresponding to individual recording channels were extracted using Cheetah 5 Data Acquisition Software (Neuralynx, Inc.). The sampling frequency remained the same (32556 Hz), while the bandpass filter frequency and the input range settings were individually tailored to each channel. Raw data were bandpass filtered with a low cut frequency of 600 Hz and a high cut frequency that ranged from 2500 to 4000 Hz, depending on the channel and session. The relatively low value for the high cut frequency was selected in order to exclude high frequency noise that was present in the data. (This noise was only detected at later stages of the analysis; if detected earlier, it could have been removed through shielding and referencing, as is now done in the lab.)

After playback, the data were not saved at 24-bit resolution, but rather at 16-bit resolution. The voltage input range was set during playback using Cheetah 5 software

(Neuralynx, Inc.), to obtain the highest resolution possible (in volts per AD count), while simultaneously ensuring that spike amplitudes did not reach saturation levels. Values of the input ranges typically spanned 15 to 150 μV. This stage of processing generated 'continuously-sampled channel' (CSC) data from the raw data, which could then be imported into Matlab using commercially-provided MEX files (Neuralynx, Inc.) or custom-written Matlab routines.

A.2 Threshold selection for spike extraction using CSC Spike Extractor

To select thresholds for spike extraction, the data were visualised using CSC Spike Extractor software (Neuralynx, Inc.). Thresholds were placed above the noise level such that low-amplitude spikes were still included, as long as their amplitude exceeded that of the noise (an example threshold level is depicted by a horizontal white line in Figure 72). Once the desired threshold values were determined, the extraction of discrete spike waveforms from the CSC data was carried out using custom-written Matlab routines.

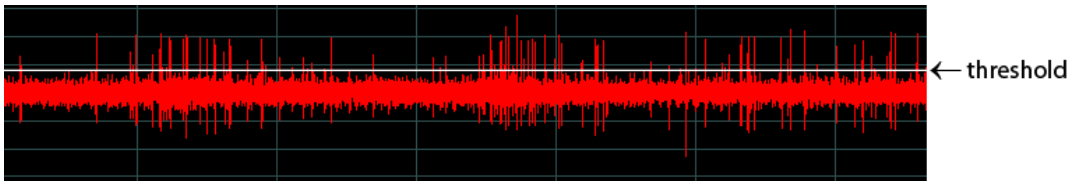


Figure 72. Thresholds were selected with the aim of maximising noise exclusion and spike inclusion (based on human judgment). The horizontal white line depicts the threshold level.

A.3 Artifact removal

The steps of threshold-setting and spike extraction described above resulted in the generation of data in the form of discrete spike waveforms, from the CSC data. SpikeSort3D software (Neuralynx, Inc.) was used to visualise these spike waveforms, as well as ISI histograms, 3D plots of waveform features, and 'time plots' depicting selected feature properties across an entire session. A visual inspection of the signals obtained across all recording sessions, during manual spike sorting, revealed that most of the channels appeared to yield single unit activity (assuming that the uniformity in

the shape of the waveforms on each channel indicated that they originated from a single neuron); on some channels, two or more units could be discerned, which could be separated based on their waveform features with varying degrees of ease, depending on the recording session; a minority of channels consistently yielded two distinct, sortable units across the vast majority of recording sessions. To minimise the introduction of human error and biases in data selection, a conservative approach was taken, in which spiking activity was pooled across all units, regardless of how many distinct waveforms were discernible on each channel. Signals obtained from each recording channel were thus deemed to be 'multiunit activity' rather than 'single unit activity.'

A.3.1 Examination of rasters across all recording sessions, for each channel

Peristimulus time histograms (PSTHs) and rasters were plotted for each channel, across all recording sessions. A close examination revealed two features that needed addressing:

1. The data were contaminated by an artifact that was generated each time the monitor refresh occurred. Its presence was unexpected as our recordings had previously been made using an analog Neuralynx system, rather than the current digital system, and such artifacts had never been detected previously in single electrode recordings. These artifacts occurred at fixed intervals across many of the channels. They varied in amplitude from day to day and from channel to channel.
2. Examination of the PSTHs revealed that for a given channel, the level of spiking activity varied considerably between days, due to variations in manual selection of thresholds (as described in the previous section).

These issues were addressed using the methods described below.

A.3.2 Artifacts induced by the monitor refresh

Monitor artifacts occurred at a fixed point in time relative to the onset of each monitor refresh. Their form was indistinguishable from the multi-unit spiking activity in

many channels and could therefore not be eliminated during the spike sorting process. However, based on the regularity of their occurrence, it was possible to calculate a 'template' of the artifact (the mean waveform) for each channel and each day. This template could then be subtracted from the raw CSC signal at the time of its occurrence. It was thus completely eliminated from the CSC (voltage) signal without inadvertent elimination of legitimate spikes.

The timing of each monitor refresh during a given trial was calculated as $t_x = t_{onset} + x\tau$, where t_{onset} is the time of stimulus onset, τ is the interval between monitor refreshes (the 'refresh interval'), and t_x is the time of a monitor refresh that occurs x refreshes away from τ. For each session, the average voltage signal was calculated across all refresh intervals and across all trials, yielding the mean signal obtained during the inter-refresh period (Figure 73). The peak in the number of spikes due to the monitor artifact occurred at around 1 ms from the start of each refresh (monkey 1, V4 location: 0.97 ms, V1 location: 0.95 ms; monkey 2, V4 location: 1.15 ms, V1 location: either 0.97 or 1.19 ms).

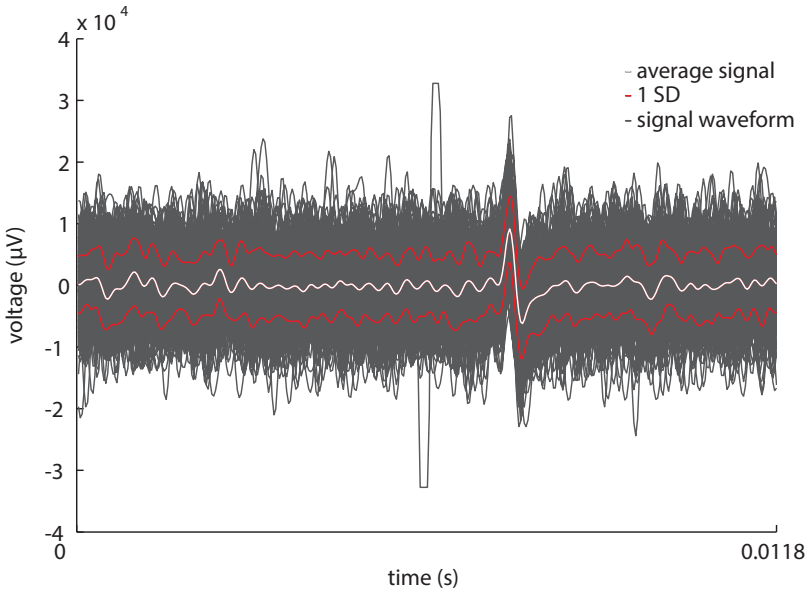


Figure 73. Waveforms recorded during all refresh intervals across both correct and incorrect trials, from a single session for an example channel (monkey 1, channel 4, session 333, V4 location), plotted on the same graph and aligned to the same point in the monitor refresh cycle (overlapping grey lines). Time = 0 corresponds to the time at which the computer issued the command for stimulus presentation on the first trial, and every subsequent time point (in multiples of the inter-refresh period) after that. The

average signal taken across all occurrences of the monitor refresh is represented by the white line; this corresponds to the waveform of the monitor-induced artifact. Red lines depict 1 SD from the mean.

A comparison of PSTHs and rasters that were generated before and after this procedure verified that the monitor-induced artifacts had been successfully removed (Figure 74).

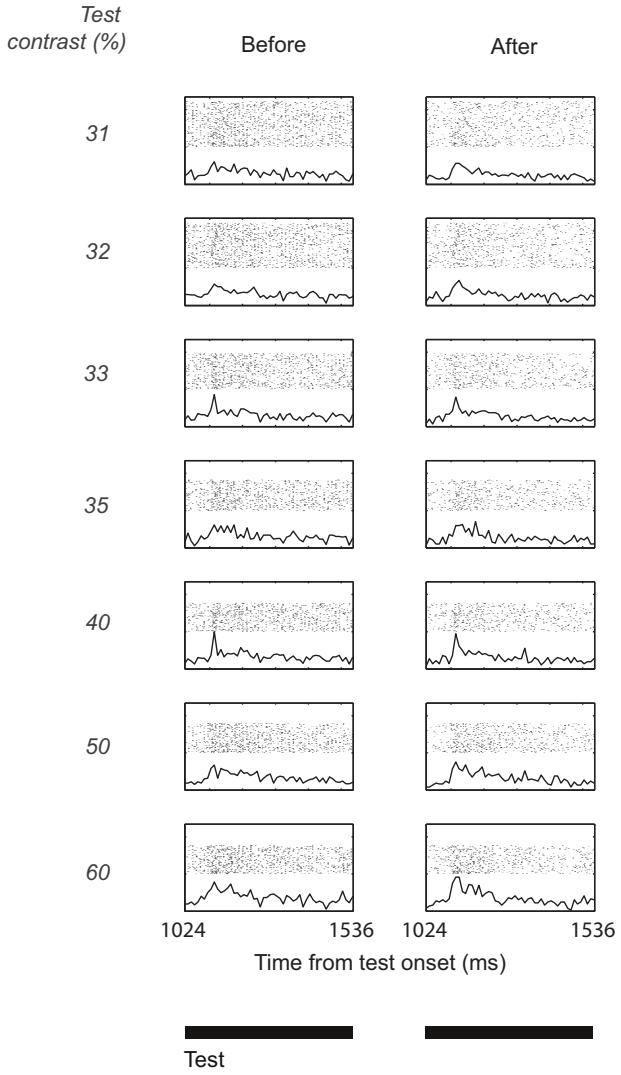


Figure 74. Rasters plotted for each trial, against time, for conditions with test contrasts of 31, 32, 33, 35, 40, 50, and 60%, during test stimulus presentation (1024 to 1536 ms relative to sample onset), from an example channel and session (monkey 1, channel 4, session 332). Left column: before artifact removal; right column: after artifact removal. Note the presence of artifacts due to each monitor refresh in the left plots- rasters are

contaminated by regularly-spaced artifacts that are temporally aligned to stimulus onset, and appear in the form of thin vertical stripes that run across trials. Artifacts also show up in the PSTHs, generating extraneous peaks at regularly-spaced intervals. Of all the channels from which recordings were made, this was one of the most badly-contaminated examples; the raster plots of most channels did not contain such clearly-visible artifacts. After artifact removal, signs of artifacts are greatly reduced or absent.

A.3.3 Automated threshold setting to obtain uniform spontaneous activity levels across sessions

Although the initial stages of threshold setting and spike extraction were conducted manually (using CSC Spike Extractor software), this method did not yield closely-matched levels of spiking activity during the spontaneous period, across sessions. As such, our next step was to ensure that spontaneous activity levels remained consistent across sessions.

For each trial, one would expect that the period during which activity levels remained minimally affected by training should be that which occurred prior to sample onset (i.e. during the spontaneous period). We were aware that it was not possible to be certain that training did not affect the pre-sample spontaneous activity; however, in comparison to other periods within the trial (the stimulus-induced response and the inter-stimulus interval), the pre-sample spontaneous period appeared to be the most suitable candidate for an across-session reference. Thus, an additional step of data processing was implemented, in which the selection of thresholds for spike extraction was automated using a Matlab routine, based on levels of spontaneous activity.

First, it was necessary to select a target level of spontaneous activity, which would be used as a reference across sessions. For each channel, raster plots and PSTHs were examined by eye, and a session which had ‘medium’ signal quality (i.e. with an ‘average’ SNR, compared to other sessions, and with satisfactory stimulus-induced responses), was selected as the reference. The level of spontaneous activity obtained during this session, r_t, was taken as the ‘target’ level across all sessions, for that particular channel. We were aware that this was an arbitrary choice; however, any choice would have been arbitrary.

Once the value of r_t, was selected, a suitable threshold had to be determined for each session such that levels of spontaneous activity, r_s, lay within 1% of the target value. Spontaneous activity levels depended on the threshold value, thus an iterative procedure was implemented in which spike extraction thresholds for the non-reference sessions were adjusted using a staircase procedure until the spontaneous firing rate of a given session deviated by no more than 1% from the target rate.

An examination of PSTHs that were generated via this procedure confirmed that the standardisation of spontaneous activity levels across sessions was carried out successfully (Figure 75). For the channel depicted in the example figure, the SD in firing rate across sessions prior to spontaneous activity matching was 4.57 spikes/s (mean = 13.74 spikes/s); after activity matching, the SD was reduced to 0.04 spikes/s (mean = 15.49 spikes/s).

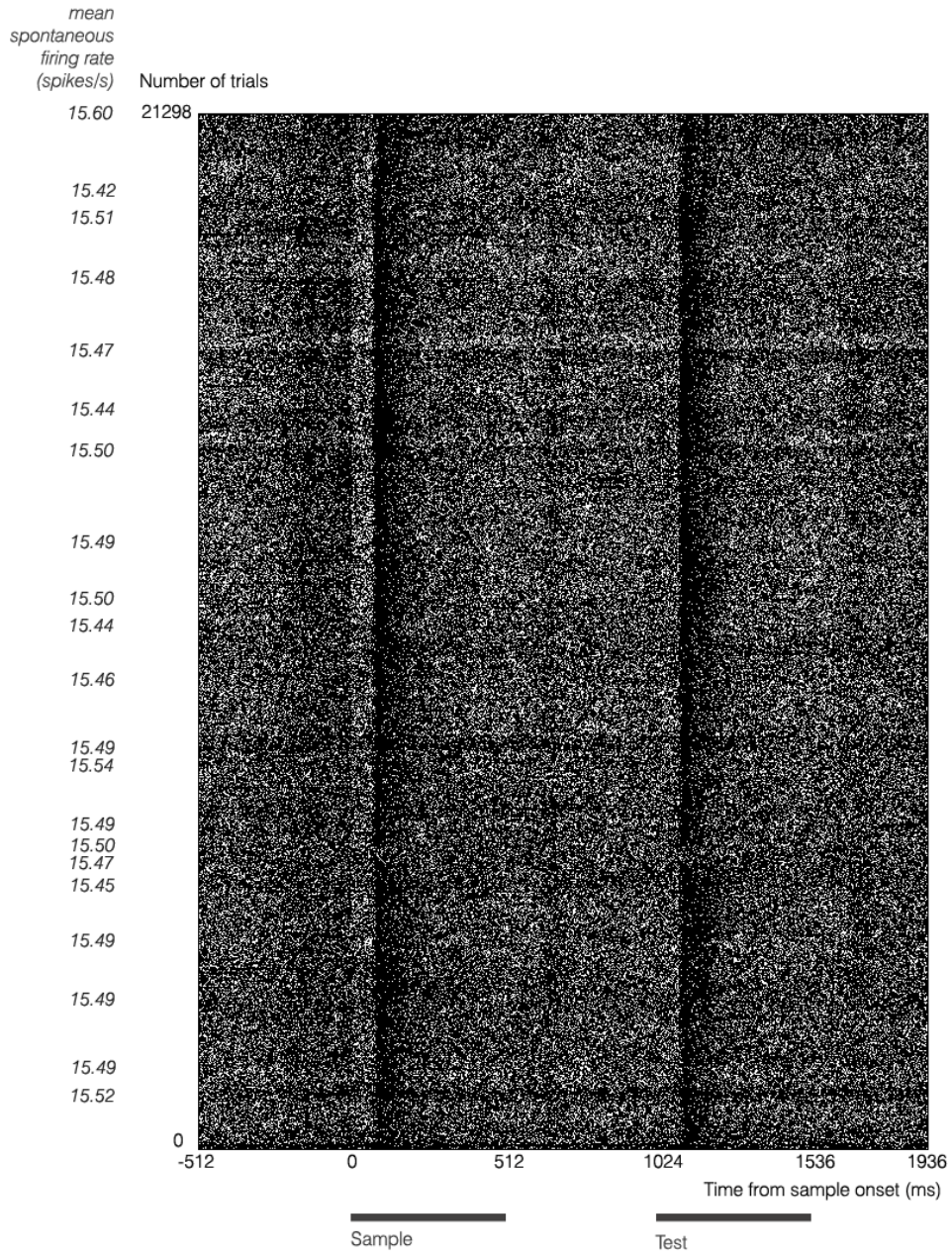


Figure 75. Rasters plotted across multiple sessions, over a total of 21,298 trials for the same channel as that shown in Figure 74 (monkey 1, channel 4). To the left of the plot, the mean spontaneous firing rate is displayed for each session. The spike extraction threshold was derived using an automated staircase procedure, and the threshold for each session was selected such that the mean spontaneous rate differed by less than 1% across sessions. Levels of neuronal activity became much more uniform across sessions, and the SD in spontaneous activity levels between sessions was markedly reduced.

A.3.4 Artifacts induced by subjects' movements

On some days of recording, physical movements by the subject resulted in the generation of high-amplitude artifacts that occurred across multiple recording channels. These movement-induced artifacts were observed during both data acquisition and data playback with Cheetah 5 software. They also showed up in the raster plots as streaks of rapidly-occurring, temporally continuous 'spikes,' at frequencies that were much higher than those of real spikes, often appearing across multiple channels (refer to Figure 76 for an example session). They typically occurred on 2 – 3% of the trials in which the subject made a correct response (monkey 1: mean = 3.14%, SD = 2.51; monkey 2: mean = 2.16%, SD = 2.35). Since these artifacts could not be cleanly removed from the rest of the signal during each trial, and they occurred on a small minority of trials per session, contaminated trials were selectively excluded, as described below.

Figure 76. Rasters and PSTHs for an example channel (monkey 2, channel 7) from a session in which movement-induced artifacts were found to occur during 57 trials (4.44% of all correct trials for that session- a particularly badly affected session). Artifacts show up in the form of semi-continuous horizontal lines which last tens of milliseconds. Trials that are contaminated by artifacts have rasters plotted in red.

To identify contaminated trials, we exploited the fact that the movement-induced artifact appeared across multiple channels simultaneously. The level of correlation in the signal between channels, for a contaminated trial, was much higher than that between channels for an uncontaminated trial. For each trial, a coefficient of correlation, *R*, was calculated for every pairwise combination of channels. The distribution of all *R*-values, obtained across all trials and pairwise comparisons, was plotted for each session. For sessions without contaminated trials, the *R*-values were distributed unimodally with a mean of around 0.2 to 0.4, depending on the session. For sessions containing contaminated trials, however, this distribution was bimodal, with a second, smaller group of *R*-values that were distributed about a higher mean that ranged from around 0.4 to 0.7 (refer to Figure 77 for the histograms obtained from two example sessions).

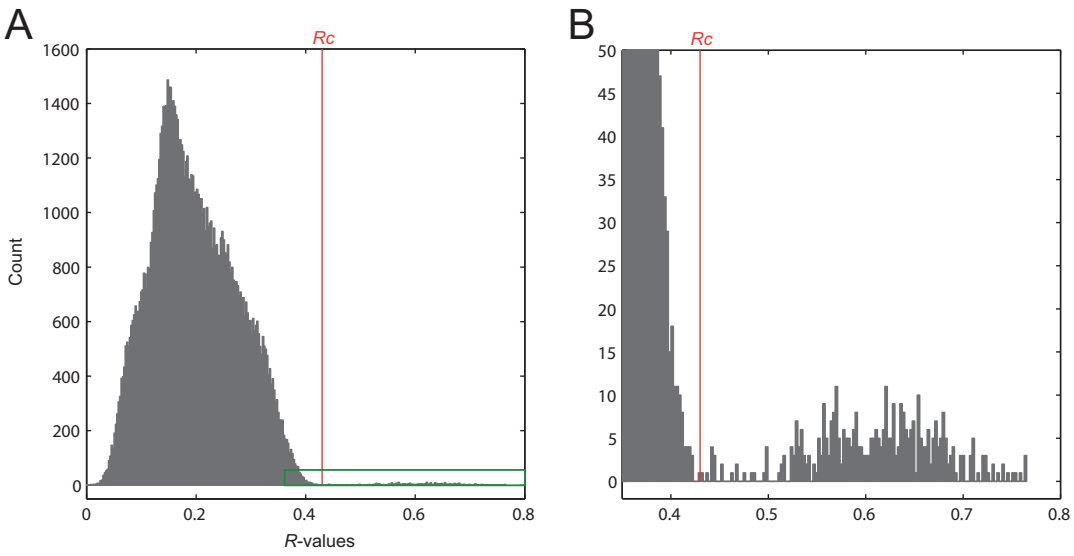


Figure 77: Histograms of *R*-values obtained from pair-wise comparisons of trials, during an example session (monkey 2, session 73). A: Histogram depicting all the *R*-values from the example session. B: Zoomed-in plot of the right tail of the histogram depicted in the left subplot (marked by the green box). Red vertical lines depict the threshold, R_c (set at 0.43 for this session).

A cut-off value, R_c, was manually chosen for each session, based on an inspection of the histogram of *R*-values, such that a maximum number of *R*-values from the higher, outlying group (if present) lay above the cut-off point. For each pairwise comparison that yielded an *R*-value higher than R_c, the trial to which it corresponded was excluded from further analysis for all channels, regardless of which pair of channels had produced that particular *R*-value. A comparison of raster plots and PSTHs, obtained

before and after exclusion of trials with movement-induced artifacts, confirmed that the unwanted trials had been successfully identified and removed (compare Figure 76 with Figure 78, for a demonstration of artifact removal for an example channel).

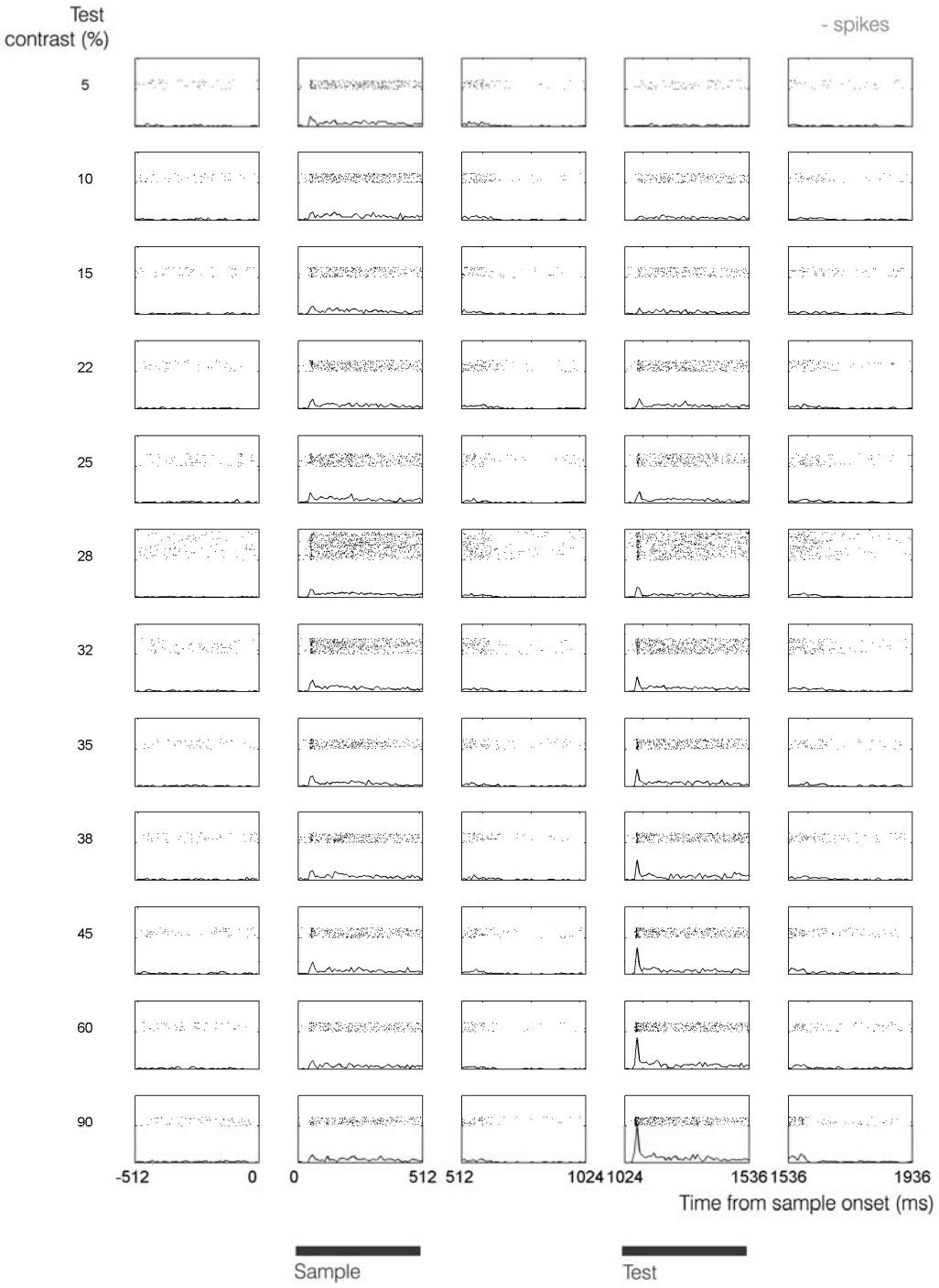


Figure 78. Rasters and PSTHs for the same channel and session as that presented in Figure 76, after the removal of trials containing movement-induced artifacts.

A.3.5 Inclusion of channels based on the signal-to-noise ratio of spiking activity

Clear stimulus-evoked activity was present on the majority of channels from which we recorded. However, the depth of our chronically-implanted electrodes could not be adjusted to maximise the quality of our recordings during each session, and the orientation and spatial frequency of our stimuli were fixed (i.e. not optimized to the preference of the neurons recorded). Thus, a few channels yielded poor data throughout most of the recording sessions, while other channels yielded good data for the majority of sessions but poor data on a few occasions. Thus, a signal-to-noise ratio (SNR) was calculated for each channel, and data from that channel were included only if the SNR exceeded a minimum value. This allowed the inclusion of a maximal amount of high-quality data, while reducing contamination due to noise.

The SNR was calculated as the ratio of the mean peak response in stimulus-evoked activity across trials (during presentation of the test) to the SD of pre-stimulus activity (during the 512 ms before sample onset) for each test contrast condition, yielding a set of fourteen SNR values per recording session for a given channel (Self, Kooijmans, Supèr, Lamme, & Roelfsema, 2012). Trials were included regardless of whether the subject's response was correct. The size of the SNR varied depending on the test contrast; since the purpose of this analysis was to determine whether the quality of the stimulus-evoked response qualified the channel for inclusion under any of the conditions used during the task, the highest of the fourteen SNR values was then taken as being representative of the signal quality from a given channel for each session.

Note that in principle, one could simply calculate the SNR for the highest test contrast, as most channels would respond best when high contrast stimuli are presented. However, this would fail to account for the (few) channels that showed non-monotonic contrast tuning. The existence of such neurons in V4 has only been documented on a few occasions in the literature- in an examination of attention effects on the CRF by Williford (2006) (refer to their Figure 5C), and in a recent publication from the Chelazzi lab (Sani, Santandrea, Golzar, Morrone, & Chelazzi, 2013). The presence of such

neurons in V1 and V2 has been reported by Peng and Van Essen (2005). Note, however, that since our electrodes recorded MUA, not single-unit activity (SUA), it is possible that the channels for which we observed non-monotonic contrast response functions were sampling from a combination of cells that collectively displayed contrast-tuned activity.

Channels with poor SNRs (less than 1) on ≥ 20% of sessions were excluded completely from further analysis. The cut-off value of 1 was chosen as it provided a maximally inclusive standard- essentially, channels were included as long as some level of stimulus-evoked activity could be detected on at least 80% of sessions. For these remaining channels, if the SNR value was less than 1 during some of the sessions (up to a maximum 20% of sessions, by definition), then only the sessions with good responses were included, while the rest were discarded. Note that this selection process resulted in the inclusion of a slightly different set of channels from one session to the next.

Appendix B: Cross correlations between PSTH waveforms of channels

Over the course of training, the implanted grid remained physically fixed in position, and recordings were made from each electrode/recording channel on each day. The question arose as to whether the identities of the neurons that were sampled by a given electrode remained largely the same throughout the training process, or whether their identities varied over time. A visual inspection of the raster plots and peristimulus time histograms (PSTHs) of visually-evoked responses revealed that on many of the channels, the stimulus-evoked responses registered on each channel tended to adopt a characteristic pattern of activity. Numerous channels, particularly in the V4 region, could be identified by eye, based on the amplitude of their response and the temporal pattern of their spontaneous and stimulus-evoked activities (refer to Figure 79 for examples of channels with distinctive firing patterns). Furthermore, the shape of the PSTH from a given channel tended to remain highly consistent over the course of training.

In the absence of microscopy and cell staining techniques, it was not possible to positively identify the neurons that were being sampled by each electrode across recording sessions; however, it was possible to continually monitor the responses obtained from each channel throughout the training period (Nicolelis et al., 2003), and to compare the shape and time course of these responses between sessions as well as between channels, thus providing a general idea of the levels of variability within the data (Dickey, Suminski, Amit, & Hatsopoulos, 2009).

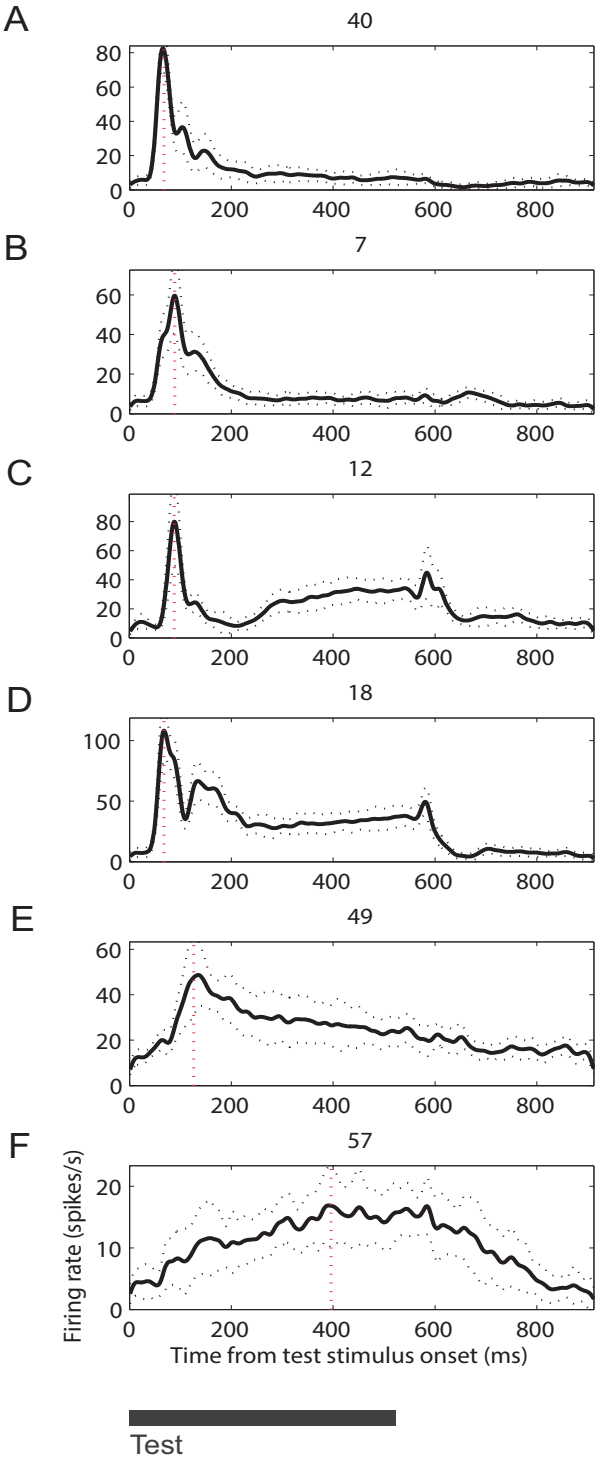
A cross-correlation analysis was performed on PSTHs from individual channels, to quantify similarities in these responses across sessions. Levels of inter-session correlations in spiking activity from a given channel were compared with those seen *across* different channels. This analysis was performed using data from sessions for which at least 30 correct trials per condition had been recorded, and it focused on the period of time during which test-stimulus-evoked spiking activity was elicited (from the onset of the test stimulus, to 400 ms after its offset, spanning 512 + 400 = 912 ms per trial). Throughout this analysis, only data from trials with correct responses were used. Generally speaking, the higher the contrast of the stimulus, the stronger the neuronal

response. Thus, this analysis included only the data obtained in response to presentations of test stimuli with the highest contrast levels (60% contrast for stimuli at the V4 location; 90% contrast for stimuli at the V1 location).

B.1 Methods

First, a bootstrap procedure was carried out on data from individual recording channels, for a given session. This provided a measure of the amount of variability that could be expected from recordings that were taken from a single channel over a short period of time (typically up to 3 or 4 hours per session). The trials obtained from a particular channel, during a particular session, formed a pool of data; the number of trials constituting this pool was designated as *n*. A set of *n* trials was randomly selected, with replacement, from the pool. A PSTH was generated across this new set of trials, using a smoothing window of 16 ms. This process of selection with replacement and PSTH generation was conducted 100 times, yielding 100 sets of bootstrapped PSTH values. To assess the levels of variability of the visually-evoked response within this bootstrapped dataset, a correlation analysis was carried out using the *xcorr* function in Matlab. A correlation coefficient value, R_b, was generated for all possible pairwise combinations of bootstrapped data ($the\ number\ of\ R_b\ values\ = \frac{100!}{2!(100-2)!} = 4950$). The mean and standard deviation of the distribution of R_b values were calculated. This stage of analysis was carried out separately for each channel and each session.

Next, PSTH values were calculated using the original, complete set of data (without carrying out bootstrapping), for each channel and each session. PSTH datasets were then pooled across sessions, for individual channels. To examine the degree of correlation between signals recorded over multiple sessions from a given channel, pairwise comparisons were carried out between pairs of PSTH datasets from multiple sessions. This yielded a set of correlation coefficients, R_a, for each channel, which described the actual variability present in the signals recorded from individual channels, across the whole training period.



A rapidly-occurring initial transient response upon stimulus onset is followed by a fairly rapid decline in activity. Response suppression occurs after the offset of the stimulus.

The peak of the initial transient response occurs slightly later than that seen with channel 40 in (A). This transient response is followed by a fairly rapid decline in activity.

The initial transient is followed by a drop in activity to near-spontaneous levels. A gradual ramping up of activity occurs, and a second transient peak is seen, following the offset of the stimulus.

The initial transient is followed shortly after by a second transient with a lower peak than the first. A sustained level of activity is maintained, and a third transient occurs after stimulus offset.

The initial peak in the response occurs later than that seen in the preceding examples. This is followed by a gradual decline in activity.

No initial transient occurs. Instead, activity builds up gradually over the course of stimulus presentation; it then declines almost as slowly.

Figure 79. Mean PSTHs across sessions for six example channels, illustrating the diversity of responses seen on individual recording electrodes, to a test stimulus of 60%

(monkey 1, V4 location). Activity was calculated by combining PSTHs across individual sessions (i.e., not the raw spike data), and taking the average. Dotted black lines indicate 1 SD from the mean. Red vertical lines demarcate the occurrence of the peak response.

Finally, to provide a measure of the collective amount of variability that was present across channels and recording sessions, sets of PSTH values that were generated from non-bootstrapped data (as described in the preceding paragraph) were pooled across all channels. Pairwise comparisons of these PSTH data yielded correlation coefficients, R_c, for the entire set of data, encompassing signals from multiple electrodes and sessions.

In summary, the distribution pattern of R_b values provided a measure of the level of variability obtained between signals from a given channel during a given session (based on the bootstrapped data). R_a gave an indication of the level of correlation between signals for a given channel, by comparing data from pairs of sessions. R_c provided an overall indication of the levels of correlation that occurred across multiple channels and sessions.

Each value of R_a and R_c (correlation coefficients that were generated from the actual data) was examined in relation to the distribution of values of R_b (generated from the bootstrapped data). If, for a given channel, values of R_a tended to fall within two standard deviations of the mean of the distribution of R_b values, whereas values of R_c tended to fall below this range, that would indicate that the visually-driven spike response tended to remain stable across recording sessions for a given channel (refer to Figure 82 for an illustration of the distribution of R-values for an example channel).

Furthermore, we predicted that the degree to which this pattern was observed would depend on the uniqueness of the visually-evoked response from a given channel: if the responses on a particular channel tended to be highly characteristic of that channel, and were simultaneously dissimilar from the responses obtained from other channels, then that channel would yield relatively unique PSTH waveforms, and thus consistently produce R_a values within the expected range. On the other hand, channels with responses that were similar in shape and temporal structure to those of some other channels would have PSTH waveforms that were harder to distinguish from the others, and would therefore have R_a values that lay *below* as well as within the expected range.

As Dickey et al. (2009) have pointed out, a stable unit can be expected to produce similar-looking responses over long periods of time, but different units may also display similar waveforms to each other.

Side note: The distribution of R_b values was negatively skewed, i.e. with a long tail on the left (refer to Figure 81, upper plot). Before confidence intervals could be calculated to describe the distribution of R_b values, the data needed to be normally distributed. Thus, prior to calculating the mean and standard deviation of the distribution, a square root transformation was performed on the data, to convert the skewed distribution into a symmetrical, unskewed one. This method allowed the preservation of information about the positions of R_b values relative to each other, while shifting the distribution as a whole, from skewed to unskewed. Tests of normality using the *lillietest* function in Matlab, as well as calculations of skewness and kurtosis of the distribution of data, were used to verify that the levels of skewness were satisfactorily reduced as a result of the transformation (refer to Figure 81, lower plot). Similarly, a square root transform was applied to the R_a and R_c values that were generated from the original, non-bootstrapped data, allowing a direct comparison between within-channel and across-channel distribution patterns. Wilcoxon rank sum tests were applied to test two hypotheses: that the median of the distribution of R_a values was no different from that of the R_b distribution; and that the median of the distribution of R_c values was lower than that of the R_b distribution.

B.2 Results

A cross correlation analysis was performed to assess whether the shape of the PSTH for a given channel remained consistent over time, and whether it uniquely identified each channel from the rest.

Bootstrapped PSTHs were generated using recorded data, and plotted along with the original (non-bootstrapped) PSTH (Figure 80). Correlations were calculated between pairs of PSTHs, yielding a population of R-values (R_b) for each channel and session (see Figure 81 for an example channel). R-values for bootstrapped data were fit with a Gaussian, and plotted together with the within-channel (R_a) and across-channel (R_c) correlation coefficients (see Figure 82 for an example channel).

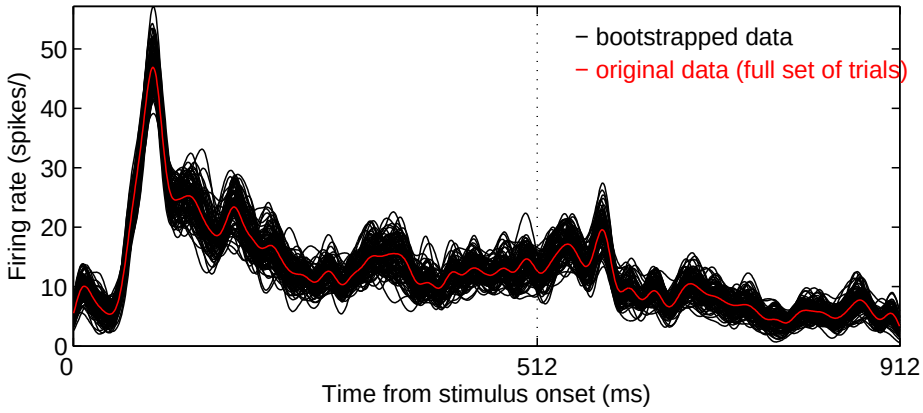


Figure 80. PSTHs generated from 100 sets of bootstrapped data (black), for an example channel and session (monkey 1, channel 7, session 336). The red line depicts the PSTH obtained from the original, full set of trials.

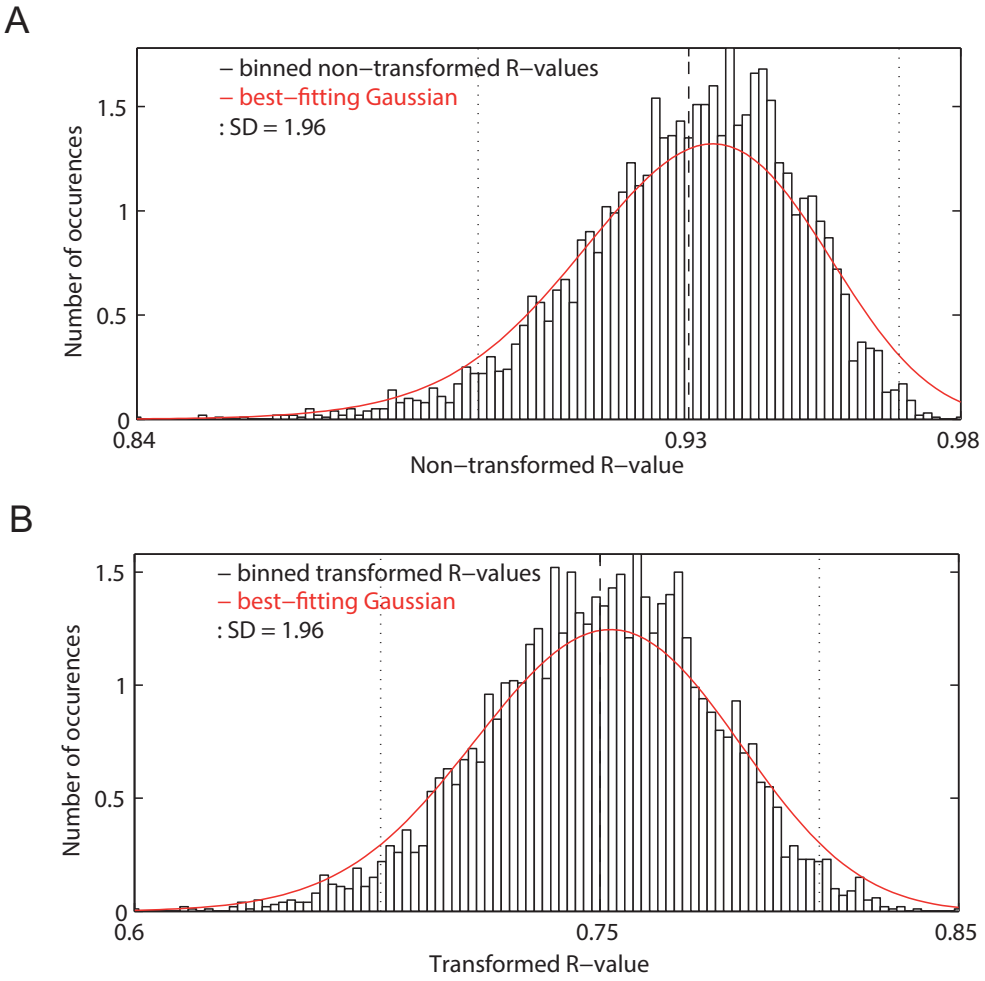


Figure 81. Histogram of R_b values for an example session (monkey 1, channel 7, session 336), before (A) and after (B) a square root transformation was applied to the data. Prior

to the transformation, the distribution was visibly skewed. Tests of skewness and kurtosis indicated that the transformation yielded a satisfactory adjustment of the data. Vertical black dotted lines indicate the mean and 1.96 SD from the mean of the distribution.

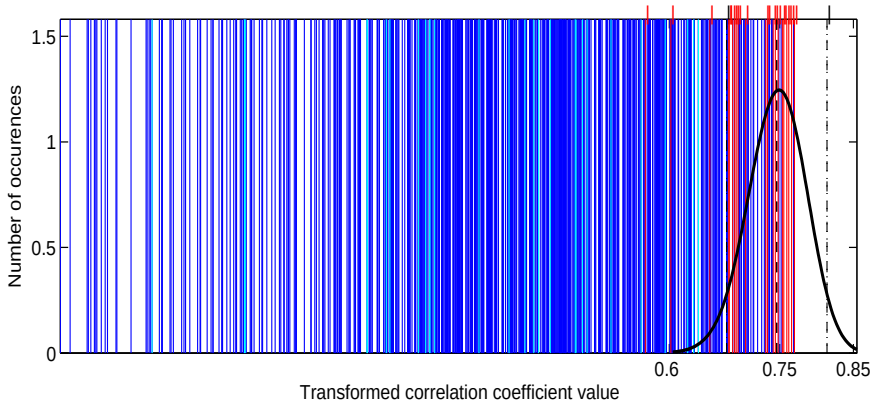


Figure 82. R_a and R_c values, in relation to the histogram of R_b values, for an example channel and session (monkey 1, channel 7, session 336). The black curve shows the best-fitting Gaussian to the distribution of R_b values. Red vertical lines depict values of R_a (within-channel, across-sessions comparisons); blue vertical lines depict values of R_c (across-channels, across-sessions comparisons). Vertical black dotted lines indicate the mean and 1.96 SD from the mean, for the distribution of R_b values. The majority of R_a values fell within the 95% interval of R_b values expected from that session, whereas the bulk of R_c values lay below this range. This indicated that out of all the PSTH responses obtained from every recording channel and every session, the responses that exhibited the greatest similarity to the one seen on that channel, on that day, tended to be those that originated from the same channel on different days.

B.2.1 Cross correlations between PSTHs of channels based on non-roving data

The 95% CI of each distribution of R_b values was determined (corresponding to 1.96 SDs below and above the mean), and the proportions of R_a and R_c values that lay within this CI were calculated for each channel. If the signal had remained relatively consistent across sessions for a given channel, then R_a/R_b would be expected to be higher than R_c/R_b. In the majority of cases, the ratio of R_a/R_b was higher than that of R_c/R_b (monkey 1, V4 location: 446/452 comparisons = 98.7%, V1 location: 235/297 = 59.2%; monkey 2, V4 location: 296/360 = 82.2%:, V1 location: 525/525 = 100.0%), indicating that the PSTH signal which was obtained from a given channel tended to remain consistent over the course of training and was largely distinct from that recorded from other channels (refer to Figure 83). Comparisons in which the value of R_a/R_b was

equal to that of R_c/R_b were excluded from this tally (monkey 1, V4 location: 12 trials excluded, V1 location: 2 trials excluded; monkey 2, V4 and V1 locations: 0 trials excluded).

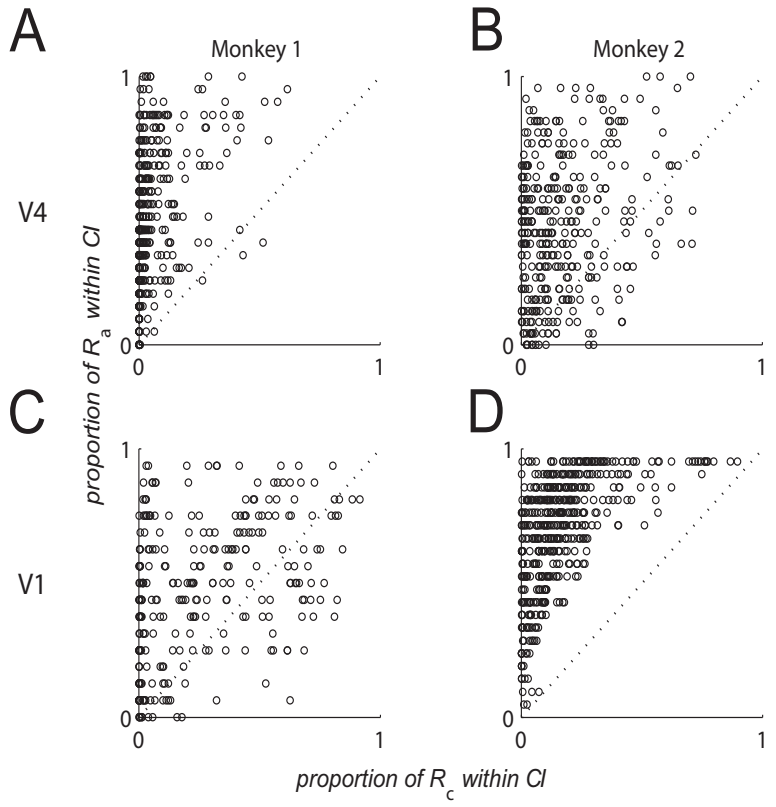


Figure 83. Scatterplots showing the proportions of R_a and R_c (y-axis and x-axis, respectively) that lay within the 95% CI of the distribution of R_b. In most cases, the proportion of R_a values that lay within the CI was higher than that of R_c values that lay within the CI, indicating that the shape of the PSTH which was obtained from a given channel tended to stay consistent over the course of training and remained largely distinct from that recorded from other channels. A & B: V4 location; C & D: V1 location. A & C: monkey 1; B & D: monkey 2.

Under ideal conditions, if each electrode recorded from a unique cortical region, and the subset of neurons sampled by each electrode remained identical from day to day, then the waveforms of activity registered by each electrode would be consistent across days (yielding a high value of $R_{a)}$. In addition, activity recorded by each electrode would be distinguishable from that seen on other electrodes (yielding a relatively low value of $R_{c)}$. The distributions of R_a and R_c would thus be distinct and non-overlapping.

In practise, two factors contributed to the existence of some degree of overlap between these distributions. Firstly, inherent variability in the signal from one day to the next resulted in reductions in R_a, even though recordings were taken from the same electrode. Secondly, when stimulus-evoked responses were highly similar across channels, this resulted in large values of R_c. Thus, as pointed out by others who have performed similar quantitative analyses of signals over long time spans (Krüger, Caruana, Dalla Volta, & Rizzolatti, 2010), our method does not offer indisputable proof that signals obtained from a given channel were always consistent or distinguishable from the others; rather, it demonstrates that levels of within-channel correlation tend to be high, and provides support for the observation that recordings were generally stable and that individual electrodes appeared to sample from more or less the same subset of neurons over time.

B.2.2 Cross correlations between PSTHs of channels based on roving data

When the same analysis was carried out on data obtained from sessions in which roving stimuli were presented, similar results were seen (refer to Figure 84). The highest possible test contrast was always 90%, regardless of the contrast of the sample, thus data from the highest test contrast condition were pooled across all three sample contrast conditions (20, 30 and 40% sample contrasts). For both subjects, the ratio of R_a/R_b was higher than that of R_c/R_b , as was seen with the analysis carried out on non-roving data (monkey 1, V1 location, roving data: 1005/1278 comparisons = 78.6%; monkey 2, V1 location, roving data: 946/949 comparisons = 99.7%). This indicated that the signal which was obtained from a given channel tended to remain consistent over the course of training and was largely distinct from that recorded from other channels. Comparisons in which the value of R_a/R_b was equal to that of R_c/R_b were excluded from this tally (monkey 1: 10 trials excluded; monkey 2: 1 trial excluded).

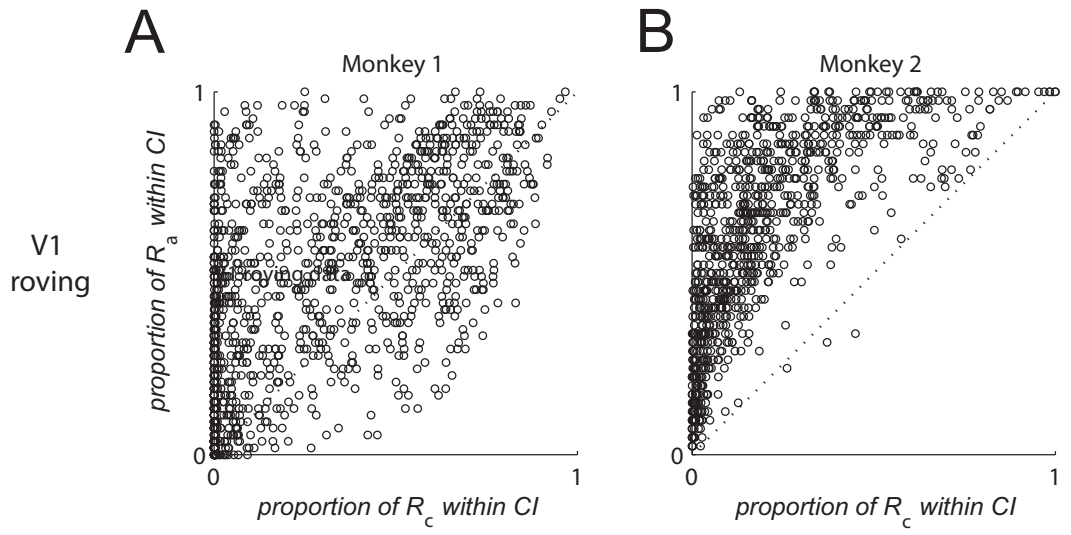


Figure 84. Results based on data collected from sessions with roving stimuli. Scatterplots show the proportions of R_a and R_c (y-axis and x-axis, respectively) that lay within the 95% CI of the distribution of R_b. A: monkey 1; B: monkey 2.

Appendix C: Characterisation of neuronal tuning properties

Analysis of the tuning properties of recorded neurons was carried out both online and offline using reverse correlation techniques that have been described elsewhere (DeAngelis et al., 1994; Gieselmann & Thiele, 2008; Ringach & Shapley, 2004), using custom-written Matlab software. Receptive field locations were mapped repeatedly over a series of recording sessions, prior to the beginning of CD task training.

C.1 Methods

Stimuli consisted of a succession of black squares (with no inter-stimulus intervals between presentations), at each of 12 possible locations on a 3-by-4 grid. The order of the grid locations at which each stimulus was displayed followed a pseudo-random sequence. The size of the squares varied over a range, which was selected based on the cortical location from which we recorded. Their edge length varied from 0.5 to 3.0 dva for recordings made at the V4 location, and from 0.1 to 3.0 dva for recordings made at the V1 location. To identify the RF location for each channel, the magnitude of the summed response to each different stimulus was calculated (normalised to spontaneous levels), to generate retinotopic maps of activity.

Orientation, phase, and SF tuning preferences were mapped on a daily basis, throughout training on the CD task. During each mapping session (held immediately prior to the onset of contrast discrimination training), the subject was presented with a continuous sequence of stimuli, consisting of a series of squarewave gratings. These gratings had one of two possible phases, and one of twelve possible orientations (evenly spaced from 0 to 165 degrees). The gratings that were used for the characterisation of tuning properties of V4 neurons were 16.0 dva in diameter, and spanned a range of either three or six spatial frequencies (monkey 1: 0.125, 0.25, 0.5, 1, 2, and 4 cycles per degree; monkey 2: 0.125, 0.25, and 0.5 cpd). Gratings at the V1 location were of 3.0 dva in diameter, and covered three spatial frequencies in both subjects (1, 3, and 7 cpd).

Levels of activity that were elicited by each combination of stimulus parameters were averaged across recording sessions. This mean activity was fit with a wrapped Gaussian, defined as

$$G(\theta) = B + A \sum_{n=-5}^{n=5} exp\left(-\frac{\left(\theta - \theta_{pref} + 180n\right)^2}{2\sigma^2}\right) \quad \ldots \text{(Equation 12)},$$

where $G(\theta)$ is the predicted response given the grating orientation (θ); A is the tuning amplitude; σ is the bandwidth; θ_{pref} is the PO (in degrees); and B is the offset, i.e. the level of spontaneous activity (Swindale, 1998; Zinke, 2006).

C.2 Results

In monkey 1, 27/29 V4 and 15/23 V1 channels displayed clear orientation tuning preferences, while in monkey 2, 20/20 V4 channels and 23/25 V1 channels displayed clear orientation tuning preferences (Figure 85). An Omnibus test for circular uniformity was used to test for uni- or multimodal deviations from uniformity in the distribution of orientation tuning preferences of neurons in each subpopulation (Berens, 2009). Channels were categorised into two groups: those with POs that lay within 45° of the vertical (45° to 135°) or of the horizontal (0° to 45° and 135° to 180°).

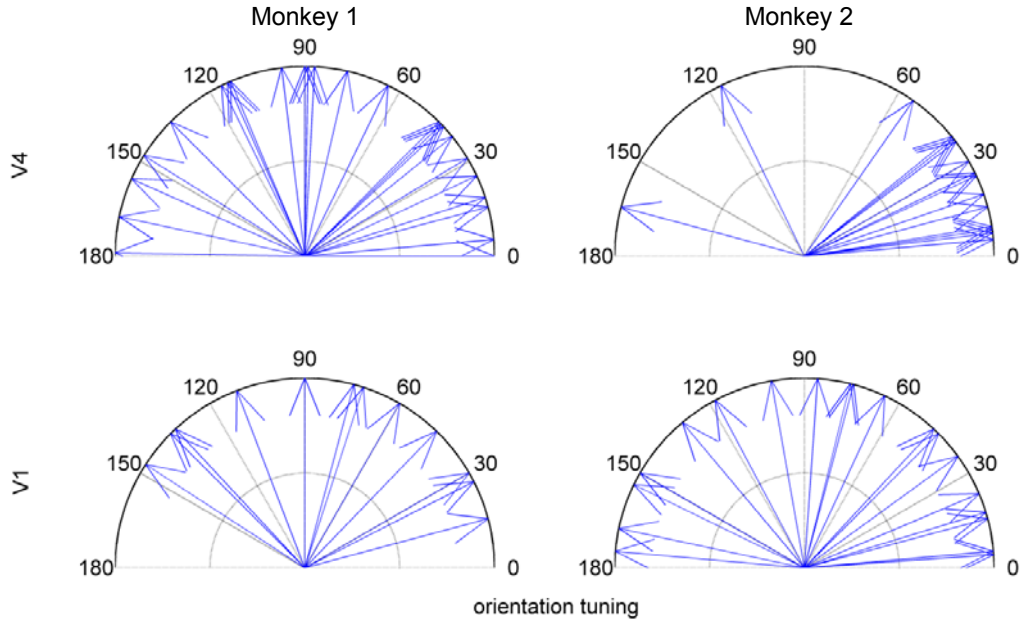


Figure 85. Distributions of orientation tuning preferences on recording channels. Left column: monkey 1; right column: monkey 2. Upper row: channels in the V4 location; lower row: channels in the V1 location.

Preferred orientations were not found to differ from a uniform distribution in three instances (monkey 1, V4: $n = 27$, $p = .629$; V1: $n = 15$, $p = .583$; monkey 2, V1: $n = 23$, $p = .265$, using a Bonferroni correction for multiple comparisons where $\alpha = .05/4 = .0125$). For recordings made at the V4 location in monkey 2, however, significantly more channels had POs that lay orthogonal to the vertical, compared to those with POs that were close to the vertical ($n = 20$, $p < .001$).

Acknowledgements

This work was supported by the Medical Research Council, UK, G0700976. Heartfelt thanks to my professor, Alexander Thiele, for invaluable instruction, supervision, and emotional support. Also to current and past members of my lab at Newcastle University: Mehdi Sanayei, Alwin Gieselmann, Mario Bartolo, Jose Herrero, David Hunter, Sascha Gotthardt, Miguel Dasilva, and Christian Brandt; and to Scott Lowe from Edinburgh University. Thanks to Jenny Read, Andrew Jackson, Stuart Baker, Anya Hurlbert, and Tom Smulders at the Institute of Neuroscience, for support and advice. Much gratitude is due to the CBC staff at Newcastle University, particularly Aurelie Thomas, Claire Richardson, Caroline Fox, Denise Reed, Ashley Waddle, Laura Watson, Mike White, and Stevie O'Keefe. Thanks to Pieter Roelfsema of the Netherlands Institute of Neuroscience for surgical expertise and assistance. Thanks also to my examiners, Adrian Rees of Newcastle University and Rufin Vogels of KU Leuven. Finally, many thanks to the board of trustees at the University of Southern California for supporting my undergraduate studies with a merit scholarship- I would not be where I am today without you.

References

Adini, Y., & Sagi, D. (2001). Recurrent networks in human visual cortex: psychophysical evidence. *Journal of the Optical Society of America A, 18*(9).

Adini, Y., Sagi, D., & Tsodyks, M. (2002). Context-enabled learning in the human visual system. *Nature, 415*(6873), 790-793.

Adini, Y., Wilkonsky, A., Haspel, R., Tsodyks, M., & Sagi, D. (2004). Perceptual learning in contrast discrimination: The effect of contrast uncertainty. *Journal of Vision, 4*(12). doi: 10.1167/4.12.2

Ahissar, M., & Hochstein, S. (1993). Attentional control of early perceptual learning. *Proceedings of the National Academy of Sciences of the United States of America, 90*(12), 5718-5722.

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences, 8*(10), 457-464

Albrecht, D. G., & Hamilton, D. B. (1982). Striate cortex of monkey and cat: contrast response function. *Journal of Neurophysiology, 48*(1), 217-237.

Anderson, J. M. (2001). Biological responses to materials. *Annual Review of Materials Research, 31*(1), 81-110. doi: doi:10.1146/annurev.matsci.31.1.81

Averbeck, B. B., Latham, P. E., & Pouget, A. (2006). Neural correlations, population coding and computation. *Nature Reviews Neuroscience, 7*(5), 358-366.

Baker, C. I., Behrmann, M., & Olson, C. R. (2002). Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. *Nature Neuroscience, 5*(11), 1210-1216.

Bao, M., Yang, L., Rios, C., He, B., & Engel, S. A. (2010). Perceptual Learning Increases the Strength of the Earliest Signals in Visual Cortex. *Journal of Neuroscience, 30*(45), 15080-15084. doi: 10.1523/jneurosci.5703-09.2010

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological), 57*(1), 289-300. doi: 10.2307/2346101

Berens, P. (2009). CircStat: A MATLAB Toolbox for Circular Statistics. *Journal of Statistical Software, 31*(10), 1-21.

Berens, P., Ecker, A. S., Cotton, R. J., Ma, W. J., Bethge, M., & Tolias, A. S. (2012). A Fast and Simple Population Code for Orientation in Primate V1. *Journal of Neuroscience, 32*(31), 10618-10626. doi: 10.1523/jneurosci.1335-12.2012

Berliner, J. E., & Durlach, N. I. (1973). Intensity perception. IV. Resolution in roving-level discrimination. *The Journal of the Acoustical Society of America, 53*(5), 1270-1287.

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *Journal of Neuroscience, 12*(12), 4745-4765.

Buracas, G. T., & Boynton, G. M. (2007). The Effect of Spatial Attention on Contrast Response Functions in Human Visual Cortex. *Journal of Neuroscience, 27*(1), 93-97. doi: 10.1523/jneurosci.3162-06.2007

Carmel, D., & Carrasco, M. (2008). Perceptual Learning and Dynamic Changes in Primary Visual Cortex *Neuron, 57*(6), 799-801.

Carrasco, M., Ling, S., & Read, S. (2004). Attention alters appearance. *Nature Neuroscience, 7*(3), 308-313.

Chen, P.-L., Chen, J.-T., Fu, J.-J., Chien, K.-H., & Lu, D.-W. (2008). A pilot study of anisometropic amblyopia improved in adults and children by perceptual learning: an alternative treatment to patching. *Ophthalmic and Physiological Optics, 28*(5), 422-428. doi: 10.1111/j.1475-1313.2008.00588.x

Chung, S., Li, R., & Levi, D. (2006). Identification of contrast-defined letters benefits from perceptual learning in adults with amblyopia. *Vision Research, 46*(22), 3853-3861. doi: http://dx.doi.org/10.1016/j.visres.2006.06.014

Chung, S., Li, R., & Levi, D. (2008). Learning to identify near-threshold luminance-defined and contrast-defined letters in observers with amblyopia. *Vision Research, 48*(27), 2739-2750. doi: http://dx.doi.org/10.1016/j.visres.2008.09.009

Cohen, M. R., & Kohn, A. (2011). Measuring and interpreting neuronal correlations. *Nature Neuroscience, 14*(7), 811-819. doi: http://www.nature.com/neuro/journal/v14/n7/abs/nn.2842.html#supplementary-information

Cohen, M. R., & Maunsell, J. H. R. (2009). Attention improves performance primarily by reducing interneuronal correlations. *Nature Neuroscience, 12*(12), 1594-1600. doi: http://www.nature.com/neuro/journal/v12/n12/suppinfo/nn.2439_S1.html

Cohen, M. R., & Maunsell, J. H. R. (2010). A Neuronal Population Measure of Attention Predicts Behavioral Performance on Individual Trials. *Journal of Neuroscience, 30*(45), 15241-15253. doi: 10.1523/jneurosci.2171-10.2010

Crist, R., Li, W., & Gilbert, C. (2001). Learning to see: experience and attention in primary visual cortex. *Nature Neuroscience, 4*(5), 519-526.

DeAngelis, G. C., Freeman, R. D., & Ohzawa, I. (1994). Length and width tuning of neurons in the cat's primary visual cortex. *Journal of Neurophysiology, 71*(1), 347-374.

Dickey, A. S., Suminski, A., Amit, Y., & Hatsopoulos, N. G. (2009). Single-Unit Stability Using Chronically Implanted Multielectrode Arrays. *Journal of Neurophysiology, 102*(2), 1331-1339. doi: 10.1152/jn.90920.2008

Dodd, J. V., Krug, K., Cumming, B. G., & Parker, A. J. (2001). Perceptually Bistable Three-Dimensional Figures Evoke High Choice Probabilities in Cortical Area MT. *Journal of Neuroscience, 21*(13), 4809-4821.

Dorais, A., & Sagi, D. (1997). Contrast Masking Effects Change with Practice. *Vision Research, 37*, 1725-1733.

Dosher, B. A., & Lu, Z.-L. (2004). Mechanisms of perceptual learning. In L. Itti, G. Rees & J. Tsotsos (Eds.), *Neurobiology of Attention*: Academic Press.

Fahle, M. (2005). Perceptual learning: specificity versus generalization *Current Opinion in Neurobiology, 15*(2), 154-160.

Fechner, G. T. (1860) *Elemente der Psychophysik* (Vol. 2). Leipzig: Breitkopf & Ha□rtel.

Fendick, M., & Westheimer, G. (1983). Effects of practice and the separation of test targets on foveal and peripheral stereoacuity. *Vision Research, 23*(2), 145-150. doi: http://dx.doi.org/10.1016/0042-6989(83)90137-2

Fine, I., & Jacobs, R. A. (2002). Comparing perceptual learning tasks: A review. *Journal of Vision, 2*(2), 190-203.

Fiorentini, A., Nicolleta Berardi. (1980). Perceptual learning specific for orientation and spatial frequency. *Nature, 287*, 43-44.

Foley, J. M., & Legge, G. E. (1981). Contrast detection and near-threshold discrimination in human vision. *Vision Research, 21*(7), 1041-1053. doi: http://dx.doi.org/10.1016/0042-6989(81)90009-2

Fredericksen, R. E., Bex, P. J., & Verstraten, F. A. J. (1997). How big is a Gabor patch, and why should we care? *The Journal of the Optical Society of America, 14*(1), 1-12.

Furmanski, C. S., Schluppeck, D., & Engel, S. A. (2004). Learning Strengthens the Response of Primary Visual Cortex to Simple Patterns. *Current Biology, 14*(7), 573-578.

Ghose, G., Yang, T., & Maunsell, J. (2002). Physiological Correlates of Perceptual Learning in Monkey V1 and V2 *Journal of Neurophysiology, 87*(4), 1867-1888.

Gieselmann, M. A., & Thiele, A. (2008). Comparison of spatial integration and surround suppression characteristics in spiking activity and the local field potential in macaque V1. *European Journal of Neuroscience, 28*, 447-459.

Gilbert, C., Sigman, M., & Crist, R. (2001). The Neural Basis of Perceptual Learning Review. *Neuron, 31*(5), 681–697.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Oxford, England: John Wiley.

Gu, Y., Liu, S., Fetsch, C., Yang, Y., Fok, S., Sunkara, A., . . . Angelaki, D. (2011). Perceptual Learning Reduces Interneuronal Correlations in Macaque Visual Cortex. *Neuron, 71*(4), 750-761. doi: 10.1016/j.neuron.2011.06.015

Hanazono, G., Tsunoda, K., Kazato, Y., Suzuki, W., & Tanifuji, M. (2012). Functional Topography of Rod and Cone Photoreceptors in Macaque Retina Determined by Retinal Densitometry. *Investigative Ophthalmology & Visual Science, 53*(6), 2796-2803. doi: 10.1167/iovs.11-9252

Herrero, Jose L., Gieselmann, Marc A., Sanayei, M., & Thiele, A. (2013). Attention-Induced Variance and Noise Correlation Reduction in Macaque V1 Is Mediated by NMDA Receptors. *Neuron, 78*(4), 729-739.

Hochstein, S., & Ahissar, M. (2002). View from the Top: Hierarchies and Reverse Hierarchies in the Visual System *Neuron, 36*(5), 791–804.

Hua, T., Bao, P., Huang, C.-B., Wang, Z., Xu, J., Zhou, Y., & Lu, Z.-L. (2010). Perceptual Learning Improves Contrast Sensitivity of V1 Neurons in Cats. *Current Biology, 20*(10), 887-894. doi: 10.1016/j.cub.2010.03.066

Huang, C.-B., Zhou, Y., & Lu, Z.-L. (2008). Broad bandwidth of perceptual learning in the visual system of adults with anisometropic amblyopia. *Proceedings of the National Academy of Sciences, 105*(10), 4068-4073. doi: 10.1073/pnas.0800824105

Jehee, J. F. M., Ling, S., Swisher, J. D., van Bergen, R. S., & Tong, F. (2012). Perceptual Learning Selectively Refines Orientation Representations in Early Visual Cortex. *Journal of Neuroscience, 32*(47), 16747-16753. doi: 10.1523/jneurosci.6112-11.2012

Kahnt, T., Grueschow, M., Speck, O., & Haynes, J.-D. (2011). Perceptual Learning and Decision-Making in Human Medial Frontal Cortex. *Neuron, 70*(3), 549-559. doi: 10.1016/j.neuron.2011.02.054

Karni, A., & Sagi, D. (1991). Where practice makes perfect: Evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Sciences, 88*, 4966-4970.

Karni, A., & Sagi, D. (1993). The time course of learning a visual skill. *Nature, 365*(6443), 250-252.

Krüger, J., Caruana, F., Dalla Volta, R., & Rizzolatti, G. (2010). Seven years of recording from monkey cortex with a chronically implanted multiple microelectrode. *Frontiers in Neuroengineering, 3*(6).

Kuai, S.-G., Zhang, J.-Y., Klein, S. A., Levi, D. M., & Yu, C. (2005). The essential role of stimulus temporal patterning in enabling perceptual learning. *Nature Neuroscience, 8*(11), 1497-1499.

Law, C.-T., & Gold, J. (2008). Neural correlates of perceptual learning in a sensorimotor, but not a sensory, cortical area. *Nature Neuroscience, 11*(4), 505-513 doi: doi:10.1038/nn2070

Lee, J., & Maunsell, J. H. (2010). The effect of attention on neuronal responses to high and low contrast stimuli. *Journal of Neurophysiology, 104*(2), 960-971.

Legge, G. E., & Foley, J. M. (1980). Contrast masking in human vision. *Journal of the Optical Society of America, 70*(12), 1458-1471.

Levi, D. (2005). Perceptual learning in adults with amblyopia: A reevaluation of critical periods in human vision. *Developmental Psychobiology, 46*(3), 222-232. doi: 10.1002/dev.20050

Levi, D., & Polat, U. (1996). Neural plasticity in adults with amblyopia. *Proceedings of the National Academy of Sciences, 93*(13), 6830-6834.

Levi, D., Polat, U., & Hu, Y. (1997). Improvement in Vernier acuity in adults with amblyopia. Practice makes better. *Investigative Ophthalmology & Visual Science, 38*(8), 1493-1510.

Li, R., Klein, S., & Levi, D. (2008). Prolonged Perceptual Learning of Positional Acuity in Adult Amblyopia: Perceptual Template Retuning Dynamics. *The Journal of Neuroscience, 28*(52), 14223-14229. doi: 10.1523/jneurosci.4271-08.2008

Li, R., & Levi, D. (2004). Characterizing the mechanisms of improvement for position discrimination in adult amblyopia. *Journal of Vision, 4*(6), 476-487. doi: 10.1167/4.6.7

Li, R., Provost, A., & Levi, D. (2007). Extended Perceptual Learning Results in Substantial Recovery of Positional Acuity and Visual Acuity in Juvenile Amblyopia. *Investigative Ophthalmology & Visual Science, 48*(11), 5046-5051. doi: 10.1167/iovs.07-0324

Li, R., Young, K., Hoenig, P., & Levi, D. (2005). Perceptual Learning Improves Visual Performance in Juvenile Amblyopia. *Investigative Ophthalmology & Visual Science, 46*(9), 3161-3168. doi: 10.1167/iovs.05-0286

Li, W., Piëch, V., & Gilbert, C. (2004). Perceptual learning and top-down influences in primary visual cortex. *Nature Neuroscience, 7*(6), 651-657. doi: doi:10.1038/nn1255

Liu, Z., & Vaina, L. M. (1998). Simultaneous learning of motion discrimination in two directions. *Cognitive Brain Research, 6*, 347-349.

Lu, Z.-L., Hua, T., Huang, C.-B., Zhou, Y., & Dosher, B. A. (2011). Visual perceptual learning. *Neurobiology of Learning and Memory, 95*(2), 145-151. doi: http://dx.doi.org/10.1016/j.nlm.2010.09.010

Martínez-Trujillo, J. C., & Treue, S. (2002). Attentional Modulation Strength in Cortical Area MT Depends on Stimulus Contrast. *Neuron, 35*(2), 365-370. doi: http://dx.doi.org/10.1016/S0896-6273(02)00778-X

Matthews, N., Liu, Z., Geesaman, B. J., & Qian, N. (1999). Perceptual learning on orientation and direction discrimination. *Vision Research, 39*(22), 3692-3701.

McAdams, C. J., & Maunsell, J. H. R. (2000). Attention to Both Space and Feature Modulates Neuronal Responses in Macaque Area V4. *Journal of Neurophysiology, 83*(3), 1751-1755.

Mitchell, J. F., Sundberg, K. A., & Reynolds, J. H. (2007). Differential Attention-Dependent Response Modulation across Cell Classes in Macaque Visual Area V4. *Neuron, 55*(1), 131-141.

Mitchell, J. F., Sundberg, K. A., & Reynolds, J. H. (2009). Spatial Attention Decorrelates Intrinsic Activity Fluctuations in Macaque Area V4. *Neuron, 63*(6), 879-888. doi: http://dx.doi.org/10.1016/j.neuron.2009.09.013

Mollon, J. D., & Danilova, M. V. (1996). Three remarks on perceptual learning. *Spatial Vision, 10*, 51-58.

Mukai, I., Kim, D., Fukunaga, M., Japee, S., Marrett, S., & Ungerleider, L. (2007). Activations in Visual and Attention-Related Areas Predict and Correlate with the Degree of Perceptual Learning. *Journal of Neuroscience, 27*(42), 11401-11411. doi: 10.1523/jneurosci.3002-07.2007

Nicolelis, M. A. L., Dimitrov, D., Carmena, J. M., Crist, R., Lehew, G., Kralik, J. D., & Wise, S. P. (2003). Chronic, multisite, multielectrode recordings in macaque monkeys. *Proceedings of the National Academy of Sciences, 100*(19), 11041-11046. doi: 10.1073/pnas.1934665100

Nienborg, H., R. Cohen, M., & Cumming, B. G. (2012). Decision-Related Activity in Sensory Neurons: Correlations Among Neurons and with Behavior. *Annual Review of Neuroscience, 35*(1), 463-483. doi: doi:10.1146/annurev-neuro-062111-150403

Otto, T. U., Herzog, M. H., Fahle, M., & Zhaoping, L. (2006). Perceptual learning with spatial uncertainties. *Vision Research, 46*(19), 3223-3233. doi: 10.1016/j.visres.2006.03.021

Palmer, C., Cheng, S.-Y., & Seidemann, E. (2007). Linking Neuronal and Behavioral Performance in a Reaction-Time Visual Detection Task. *Journal of Neuroscience, 27*(30), 8122-8137. doi: 10.1523/jneurosci.1940-07.2007

Parkosadze, K., Otto, T. U., Malania, M., Kezeli, A., & Herzog, M. H. (2008). Perceptual learning of bisection stimuli under roving: Slow and largely specific. *Journal of Vision, 8*(1), 1-8.

Pelli, D., Zhang L. (1991). Accurate control of contrast on microcomputer displays. *Vision Research, 31*(7-8), 1337-1350.

Peng, X., & Van Essen, D. C. (2005). Peaked Encoding of Relative Luminance in Macaque Areas V1 and V2. *Journal of Neurophysiology, 93*(3), 1620-1632. doi: 10.1152/jn.00793.2004

Phan, M., & Ni, R. (2011). Training Older Adults to Improve Their Contrast Sensitivity: A Possible or Impossible Task? *Journal of Vision, 11*(11), 4. doi: 10.1167/11.11.1027

Polat, U. (1999). Functional architecture of long-range perceptual interactions. *Spatial Vision, 12*(2), 143-162. doi: 10.1163/156856899x00094

Polat, U., Ma-Naim, T., Belkin, M., & Sagi, D. (2004). Improving vision in adult amblyopia by perceptual learning. *Proceedings of the National Academy of Sciences of the United States of America, 101*(17), 6692-6697. doi: 10.1073/pnas.0401200101

Polat, U., Ma-Naim, T., & Spierer, A. (2009). Treatment of children with amblyopia by perceptual learning. *Vision Research, 49*(21), 2599-2603. doi: 10.1016/j.visres.2009.07.008

Polat, U., Mizobe, K., Pettet, M. W., Kasamatsu, T., & Norcia, A. M. (1998). Collinear stimuli regulate visual responses depending on cell's contrast threshold. *Nature, 391*(6667), 580-584.

Polat, U., & Sagi, D. (1993). Lateral interactions between spatial channels: Suppression and facilitation revealed by lateral masking experiments. *Vision Research, 33*(7), 993-999. doi: 10.1016/0042-6989(93)90081-7

Polikov, V. S., Tresco, P. A., & Reichert, W. M. (2005). Response of brain tissue to chronically implanted neural electrodes. *Journal of Neuroscience Methods, 148*(1), 1-18.

Polley, D. B., Elizabeth E. Steinberg, and Michael M. Merzenich. (2006). Selective Cognitive Impairments Associated with NMDA Receptor Blockade in Humans. *Journal of Neuroscience, 26*(18), 4970–4982.

Pooresmaeili, A., Poort, J., Thiele, A., & Roelfsema, P. R. (2010). Separable Codes for Attention and Luminance Contrast in the Primary Visual Cortex. *Journal of Neuroscience, 30*(38), 12701-12711. doi: 10.1523/jneurosci.1388-10.2010

Pouget, A., Dayan, P., & Zemel, R. (2000). Information processing with population codes. *Nature Reviews Neuroscience, 1*(2), 125-132.

Raiguel, S., Rufin Vogels, Santosh G. Mysore, Guy A. Orban. (2006). Learning to See the Difference Specifically Alters the Most Informative V4 Neurons. *Journal of Neuroscience, 26*(24), 6589–6602.

Rainer, G., Lee, H., & Logothetis, N. K. (2004). The Effect of Learning on the Function of Monkey Extrastriate Visual Cortex. *PLoS Biology, 2*(2), E44.

Ramachandran, V. S., & Braddick, O. (1973). Orientation-specific learning in stereopsis. *Perception, 2*(3), 371-376.

Reynolds, J. H., Tatiana Pasternak, Robert Desimone. (2000). Attention Increases Sensitivity of V4 Neurons. *Neuron, 26*(3), 703–714.

Ringach, D., & Shapley, R. (2004). Reverse correlation in neurophysiology. *Cognitive Science, 28*(2), 147-166. doi: 10.1207/s15516709cog2802_2

Rousche, P. J., & Normann, R. A. (1998). Chronic recording capability of the Utah Intracortical Electrode Array in cat sensory cortex. *Journal of Neuroscience Methods, 82*(1), 1-15.

Saarela, T. P., Sayim, B., Westheimer, G., & Herzog, M. H. (2009). Global stimulus configuration modulates crowding. *Journal of Vision, 9*(2). doi: 10.1167/9.2.5

Saffell, T., & Matthews, N. (2003). Task-specific perceptual learning on speed and direction discrimination. *Vision Research, 43*(12), 1365-1374.

Sanayei, M. (2013). *Neurophysiology and Neuropharmacology of Perceptual Learning.* (PhD), Newcastle University, UK.

Sani, I., Santandrea, E., Golzar, A., Morrone, M. C., & Chelazzi, L. (2013). Selective Tuning for Contrast in Macaque Area V4. *The Journal of Neuroscience, 33*(47), 18583-18596. doi: 10.1523/jneurosci.3465-13.2013

Sceniak, M. P., Michael J. Hawken, Robert Shapley. (2001). Visual Spatial Characterization of Macaque V1 Neurons. *Journal of Neurophysiology, 85*(5), 1873-1887.

Schoups, A., Vogels, R., Qian, N., & Orban, G. (2001). Practising orientation identification improves orientation coding in V1 neurons. *Nature, 412*(6846), 549-553 doi: doi:10.1038/35087601

Schwartz, S., Maquet, P., & Frith, C. (2002). Neural correlates of perceptual learning: A functional MRI study of visual texture discrimination. *Proceedings of the National Academy of Sciences of the United States of America, 99*(26), 17137-17142. doi: 10.1073/pnas.242414599

Sclar, G., Maunsell, J., & Lennie, P. (1990). Coding of image contrast in central visual pathways of the macaque monkey. *Vision Research, 30*(1), 1-10.

Seitz, A. R., Nanez, J. E., Holloway, S. R., Koyama, S., & Watanabe, T. (2005). Seeing what is not there shows the costs of perceptual learning. *Proceedings of the National Academy of Sciences of the United States of America, 102*(25), 9080-9085. doi: 10.1073/pnas.0501026102

Seitz, A. R., & Watanabe, T. (2003). Psychophysics: Is subliminal learning really passive? *Nature, 422*(6927), 36-36.

Self, M. W., Kooijmans, R. N., Supèr, H., Lamme, V. A., & Roelfsema, P. R. (2012). Different glutamate receptors convey feedforward and recurrent processing in macaque V1. *Proceedings of the National Academy of Sciences, 109*(27), 11031-11036. doi: 10.1073/pnas.1119527109

Shadlen, M. N., & Newsome, W. T. (1996). Motion perception: seeing and deciding. *Proceedings of the National Academy of Sciences, 93*(2), 628-633.

Shapley, R., Michael Hawken, and Dario L. Ringach. (2003). Dynamics of Orientation Selectivity Review in the Primary Visual Cortex and the Importance of Cortical Inhibition. *Neuron, 38*(5), 689–699.

Shiozaki, H. M., Tanabe, S., Doi, T., & Fujita, I. (2012). Neural Activity in Cortical Area V4 Underlies Fine Disparity Discrimination. *Journal of Neuroscience, 32*(11), 3830-3841. doi: 10.1523/jneurosci.5083-11.2012

Simeral, J. D., Kim, S.-P., Black, M. J., Donoghue, J. P., & Hochberg, L. R. (2011). Neural control of cursor trajectory and click by a human with tetraplegia 1000 days after implant of an intracortical microelectrode array. *Journal of Neural Engineering, 8*(2), 025027.

Sowden, P., Rose, D., & Davies, I. (2002). Perceptual learning of luminance contrast detection: specific for spatial frequency and retinal location but not orientation. *Vision Research, 42*(10), 1249–1258.

Swindale, N. V. (1998). Orientation tuning curves: empirical description and estimation of parameters. *Biological Cybernetics, 78*(1), 45-56. doi: 10.1007/s004220050411

Tartaglia, E. M., Bamert, L., Mast, F. W., & Herzog, M. H. (2009). Human Perceptual Learning by Mental Imagery. *Current Biology, 19*(24), 2081-2085. doi: http://dx.doi.org/10.1016/j.cub.2009.10.060

Thiele, A. (2004). Perceptual Learning: Is V1 up to the Task? . *Current Biology, 14*(16), 671-673.

Thiele, A., Delicato, L., Roberts, M., & Gieselmann, A. (2006). A novel electrode–pipette design for simultaneous recording of extracellular spikes and iontophoretic drug application in awake behaving monkeys. *Journal of Neuroscience Methods, 158*(2), 207–211.

Thiele, A., Dobkins, K., & Albright, T. (2000). Neural Correlates of Contrast Detection at Threshold. *Neuron, 26*(3), 715–724.

Thiele, A., Pooresmaeili, A., Delicato, L. S., Herrero, J. L., & Roelfsema, P. R. (2009). Additive Effects of Attention and Stimulus Contrast in Primary Visual Cortex. *Cerebral Cortex*, 2970-2981. doi: 10.1093/cercor/bhp070

Treue, S., & Maunsell, J. H. R. (1999). Effects of Attention on the Processing of Motion in Macaque Middle Temporal and Medial Superior Temporal Visual Cortical Areas. *Journal of Neuroscience, 19*(17), 7591-7602.

Treue, S., & Trujillo, J. C. M. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature, 399*(6736), 575-579.

Tsodyks, M., Adini, Y., & Sagi, D. (2004). Associative learning in early vision. *Neural Networks 17*(5-6), 823–832.

Uka, T., Sasaki, R., & Kumano, H. (2012). Change in Choice-Related Response Modulation in Area MT during Learning of a Depth-Discrimination Task is Consistent with Task Learning. *The Journal of Neuroscience, 32*(40), 13689-13700. doi: 10.1523/jneurosci.4406-10.2012

Wang, R., Zhang, J.-Y., Klein, S. A., Levi, D. M., & Yu, C. (2012). Task relevancy and demand modulate double-training enabled transfer of perceptual learning. *Vision Research, 61*(0), 33-38. doi: http://dx.doi.org/10.1016/j.visres.2011.07.019

Watanabe, T., Jose E. Nanez & Yuka Sasaki. (2001). Perceptual learning without perception. *Nature, 413*(6858), 844-848 doi: doi:10.1038/35101601

Weber, E. H. (1850). Der Tastsinn und das Gemeingefu□hl. In R. Wagner (Ed.), *In Handwo□rterbuch der Physiologie* (Vol. 3, pp. 481–588). Braunschweig, Germany: Vieweg.

Westheimer, G. (1996). Location and Line Orientation as Distinguishable Primitives in Spatial Vision. *Proceedings: Biological Sciences, 263*(1369), 503-508. doi: 10.2307/50736

Williford, T., John H. R. Maunsell (2006). Effects of Spatial Attention on Contrast Response Functions in Macaque Area V4. *Journal of Neurophysiology, 96*(1), 40-54. doi: doi:10.1152/jn.01207.2005

Wilson, H. (1980). A transducer function for threshold and suprathreshold human vision. *Biological Cybernetics, 38*(3), 171-178. doi: 10.1007/bf00337406

Xiao, L.-Q., Zhang, J.-Y., Wang, R., Klein, S., Levi, D., & Yu, C. (2008). Complete Transfer of Perceptual Learning across Retinal Locations Enabled by Double Training. *Current Biology, 18*, 1-5. doi: DOI 10.1016/j.cub.2008.10.030

Yang, T., & Maunsell, J. H. R. (2004). The Effect of Perceptual Learning on Neuronal Responses in Monkey Visual Area V4. *Journal of Neuroscience, 24*(7), 1617–1626.

Yoshimura, Y., Dantzker, J. L. M., & Callaway, E. M. (2005). Excitatory cortical neurons form fine-scale functional networks. *Nature, 433*(7028), 868-873.

Yotsumoto, Y., Sasaki, Y., Chan, P., Vasios, C. E., Bonmassar, G., Ito, N., . . . Watanabe, T. (2009). Location-Specific Cortical Activation Changes during Sleep after Training for Perceptual Learning. *Current Biology, 19*(15), 1278-1282.

Yotsumoto, Y., & Watanabe, T. (2008). Defining a Link between Perceptual Learning and Attention. *PLoS Biology, 6*(8). doi: e221 doi:10.1371/journal.pbio.0060221

Yotsumoto, Y., Watanabe, T., & Sasaki, Y. (2008). Different Dynamics of Performance and Brain Activation in the Time Course of Perceptual Learning. *Neuron, 57*(6), 827–833.

Yu, C., Klein, S., & Levi, D. (2004). Perceptual learning in contrast discrimination and the (minimal) role of context. *Journal of Vision, 4*(3), 169-182.

Yu, C., Klein, S. A., & Levi, D. M. (2002). Facilitation of contrast detection by cross-oriented surround stimuli and its psychophysical mechanisms. *Journal of Vision, 2*(3). doi: 10.1167/2.3.4

Zanker, J. M. (1999). Perceptual learning in primary and secondary motion vision. *Vision Research, 39:7*, 1293-1304

Zenger, B., & Sagi, D. O. V. (1996). Isolating Excitatory and Inhibitory Nonlinear Spatial Interactions Involved in Contrast Detection. *Vision Research, 36*(16), 2497-2513. doi: http://dx.doi.org/10.1016/0042-6989(95)00303-7

Zhang, J.-Y., Kuai, S.-G., Xiao, L.-Q., Klein, S. A., Levi, D. M., & Yu, C. (2008). Stimulus Coding Rules for Perceptual Learning. *PLoS Biology, 6*(8), e197. doi: 10.1371/journal.pbio.0060197

Zhang, J.-Y., Zhang, G.-L., Xiao, L.-Q., Klein, S., Levi, D., & Yu, C. (2010). Rule-Based Learning Explains Visual Perceptual Learning and Its Specificity and Transfer. *Journal of Neuroscience, 30*(37), 12323-12328. doi: 10.1523/jneurosci.0704-10.2010

Zhang, T., Xiao, L.-Q., Klein, S. A., Levi, D. M., & Yu, C. (2010). Decoupling location specificity from perceptual learning of orientation discrimination. *Vision Research, 50*(4), 368-374. doi: 10.1016/j.visres.2009.08.024

Zhou, Y., Huang, C., Xu, P., Tao, L., Qiu, Z., Li, X., & Lu, Z.-L. (2006). Perceptual learning improves contrast sensitivity and visual acuity in adults with anisometropic amblyopia. *Vision Research, 46*(5), 739-750. doi: http://dx.doi.org/10.1016/j.visres.2005.07.031

Zinke, W., M. J. Roberts, K. Guo, J. S. McDonald, R. Robertson, A. Thiele. (2006). Cholinergic modulation of response properties and orientation tuning of neurons in primary visual cortex of anaesthetized Marmoset monkeys. *European Journal of Neuroscience, 24*(1), 314–328. doi: Cholinergic modulation of response properties and

Zivari Adab, H., & Vogels, R. (2011). Practicing Coarse Orientation Discrimination Improves Orientation Signals in Macaque Cortical Area V4. *Current Biology, 21*(19), 1661-1666.