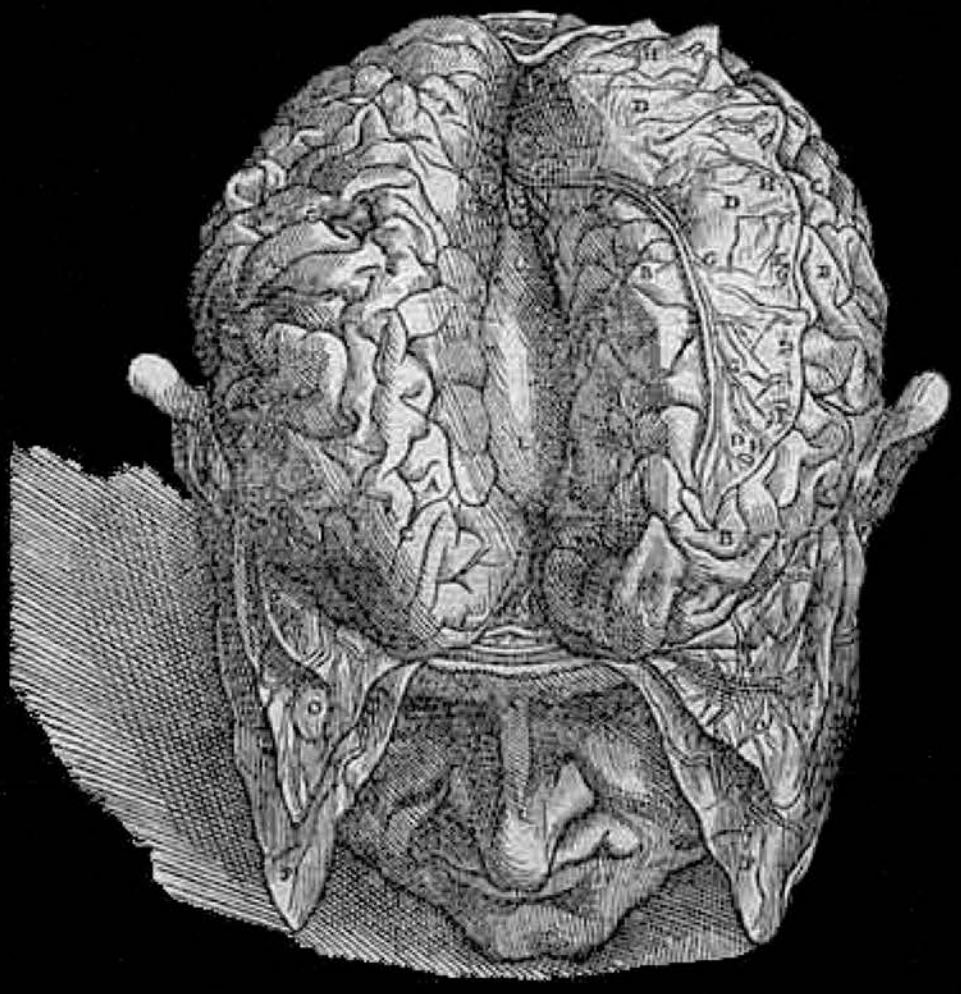
Christian KEYSERS

THE SPEED OF SIGHT

Neural correlates of perceptual reports in RSVP



Thesis submitted in June 2000 to the University of St Andrews for the degree of Doctor of Philosophy

Declarations:

I, Christian Keysers, hereby certify that this thesis, which is approximately 85000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree

Date..1/9/00.....signature of candidate....

I was admitted as a research student in August 1997 and as a candidate for the degree of Doctor of Philosophy in October 1998; the higher study for which this is a record was carried out in the University of St Andrews between 1997 and 2000.

Date.1/9/00.....signature of candidate...

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date.20/9/2000signature of supervisor... .....

Thesis copyright:

Unrestricted

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker.

Date...1/9/00.....signature of candidate... ..

To my memetic brother and friend Bas, in thankfulness for the best friendship a man can wish, both personally and scientifically. It is to our friendship, to our long and passionate discussions about science and philosophy – often conducted around a good bottle of Haut-Médoc Cru Bourgeois – that I owe the choice of going into research.

Acknowledgements

Neurophysiological investigations, such as the ones of this thesis, are the product of a team effort. With Dengke Xiao, David Perrett and Peter Földiák I really understood what a team effort really means.

Firstly, the research project described in this thesis was based on a project investigating if RSVP could be used to test neurones with thousands of rapidly presented images to find relatively 'optimal' stimuli for single neurones. Starting from that project, developed by David Perrett and Peter Foldiak and implemented by Dengke Xiao, I joined the team and found interest in testing the relationship between perception and presentation rate in RSVP. This thesis would hence not have been possible if it were not for the pre-existing project on RSVP.

Through his talented programming and his 'lets do it' attitude, Dengke solved the many technical challenges of using RSVP presentation with monkeys. His humour, his skills in single cell physiology and his positive attitude made the hours we spent together recording a pleasure. I will miss working with Dengke.

When I decided to come to St Andrews for my PhD, I had placed great expectations in working with David Perrett. He struck me, from our very first meeting, as one of the rare people who can truly listen to new ideas: he sees their value and their weaknesses with splendid clarity. With a few lucid questions he always manages to sets me on a journey to improve my own ideas. Unlike most scientists, during our discussions, he gave me the feeling of caring more about my ideas than about his own. In German, a Ph.D. supervisor is called a 'PhD-Father': that is exactly what Dave has been for me: a father helping me to mature scientifically.

I'm very thankful to Peter also, for the valuable advice he gave me throughout the thesis, especially about information theoretical analysis: he gave me new perspectives on how to interpret single cell data.

In addition to these members of the research team directly involved in the work, I would like to thank Antonella, my partner, for her support and patients throughout this thesis and for her understanding of the fact that my work forced us apart during many often difficult month. Thank you also to Bas, who gave me critical advice on many elements of my research. Throughout our many years of friendship,

he has sharpened my way to think about scientific problems, and when it came to choose my place of PhD research, he encouraged me to choose Dave's lab, and thereby helped me to take a right decision.. I would not quite be who I am, if we had not known each other for so long. Thank you to my parents, who gave me the best support I could have hoped for throughout my education Thank you also to Ruth Bennett and Juan Delius, two researchers who inspired me to pursue research in neuroscience.

My work would not have been possible without the excellent technical support from Heather Bowman, Andy Burnley, Brian Kirk, Mary Latimer, Malcolm McCandless, Wendy Taylor and Pete Wilcox.

I would like to thank all the other people around me in the university who helped make my PhD such an enjoyable experience: Chris Baker, Michael Burt, Tjeerd Jellema, Tony Little, Erika Lorincz, Mike Oram, Bernard Tiddeman and Bruno Wicker.

Finally, I am especially indebted to the Boehringer Ingelheim Foundation, who financed me during my PhD. Not only was their stipend generous, but the unbureaucratic and person way in which they supported me throughout the thesis, allowing me to attend impressive conferences, was truly an experience to remember. The research itself was financed by a research grant of the BBSRC.

Abstract

Macaque monkeys were presented with continuous rapid serial visual presentation (RSVP) sequences of unrelated naturalistic images at rates of 14ms/image to 222ms/image, while neurones that responded selectively to complex patterns (e.g. faces) were recorded in the anterior superior temporal sulcus (STSa). Stimulus selectivity was preserved for 65% of these neurones even at surprisingly fast presentation rates (14ms/image=72images/s). Such rapid processing constrains theories of visual processing. Five human subjects were asked to detect or remember specific images in the RSVP sequences under equivalent conditions. Their performance in both tasks was above chance at all rates (14-111ms/image). The neurometric performance of single neurones was quantitatively comparable to the psychophysical performance of human observers and responded in a similar way to changes in presentation rate.

Large sections (51 or 93ms) of the stimulus presentation duration in the RSVP sequences were then replaced with gaps of blank screen. This manipulation affected neither the neurometric performance of single neurones nor the recognition performance of human observers. This indicates that in STSa neural persistence after a stimulus has been turned off is quantitatively identical to the response that occurs if the stimulus stays on for up to 93ms longer: a neural correlate of human visual persistence. This maintained performance in judging the identity of a stimulus is surprising considering how different sequences with and without gaps appeared to the observers: introducing gaps as short as 23ms consistently created a perception of flicker.

Together these findings indicate that the perception of stimulus identity is dissociated from other aspects of perception such as flicker, and that the responses of STSa cells are a neural correlate of visual identity perception but not of other aspects of visual perception such as flicker.

Table of content

1. Introduction

The brain is an immensely complex organ: 10^{12} neurones interact through 10^{14} synaptic connections. Ionic flow though lipid bi-membranes creates action potentials. After decades of research, we still have no understanding of how this machinery creates the most omnipresent of all phenomena: our mind.

In the last decades however, an understanding of the mechanisms underlying brain functions has started to emerge. Particularly in the visual system, we have an increasing knowledge of the basic functional architecture of the nervous system. Most of this knowledge has been gathered through investigating the visual system of cats and macaque monkeys as models of the human visual system. Thirty two visual areas have been identified in the brain of macaques, and the connection patterns between these areas are increasingly well known (Felleman and Van Essen, 1991; Jouve et al., 1998; Young, 1992). The aim of this thesis will be to contribute to our understanding of the functioning of the brain by investigating the response of single neurones under strong time constraints and to use that knowledge to shed some light on the questions of how brain activity relates to the mind.

1.1. The distributed hierarchical model of visual processing

A basic 'text book' account has emerged from the last decades of visual neurophysiology: the distributed hierarchical model. This framework will be the basis of the investigations of the present thesis. According to this model, two main visual streams run in parallel, processing different attributes of the visual world (Ungerleider and Mishkin, 1982; Mishkin *et al.*, 1983, Goodale and Milner, 1992; Milner and Goodale, 1993): the dorsal visual stream and the ventral visual stream. The dorsal visual stream processes space and motion while the ventral stream processes shape, colour and identity. Within each stream, the properties of single neurones become increasingly complex, as one moves further away[1] from the retina (e.g. Kobatake and Tanaka, 1994). Each area uses the output of the previous area to extract increasingly complex visual features from the retinal input. In the ventral stream, LGN neurones

[1] A neurone is said to be 'further away' from the retina, if more synapses occur between the retina and this neurone.

detect[2] dots, V1 neurones detect oriented bars and so on, through V2, V4, posterior IT (inferior temporal), central IT, anterior IT and finally the anterior superior temporal sulcus (STSa), where neurones appear to detect complex, often biologically relevant objects and events, such as faces (e.g. Bruce *et al.*, 1981; Perrett *et al.*, 1982; Desimone *et al.*, 1984; Rolls, 1984; Baylis *et al.*, 1985; Yamane *et al.*, 1988; Oram and Perrett, 1992; Perrett *et al.*, 1992). The increasing complexity of the stimuli required to activate neurones as one records further away from the retina is thought to arise primarily through feed-forward convergence: the input from a number of LGN cells with collinear receptive fields is for instance supposed to create a bar-detector in the next visual area (V1) simply by converging their synaptic input onto that V1 neurone.

This 'text book' model is, like every text book model, a gross over-simplification of reality. It is important to be aware of its limitations. The principle of feed-forward processing, where information flows primarily from earlier areas ('closer' to the retina) to later areas (further away from the retina), has been challenged on the grounds that roughly equal numbers of synaptic connections flow the other way around: from the 'later' to the 'earlier' visual areas (e.g. Hupé et al., 1998; Felleman and Van Essen, 1991, see in particular Bullier and Nowak, 1995, for a critical discussion of the arguments contradicting the hierarchical model). Yet, the increasing complexity of stimuli required to activate the neurones along the ventral stream (Kobatake and Tanaka, 1994) and the roughly increasing visual response latencies along the stream (e.g. Schmolesky et al., 1998) support the idea of basically feed-forward visual processing, albeit with shortcut routes, where information can reach very high areas surprisingly quickly.

Also the idea of two separate visual streams has been challenged by an increasing amount of evidence. First, position in space has been shown to be represented also in visual areas thought to belong to the ventral stream (e.g. Baker, Keysers et al., 2000). Second, the shape of objects has been shown to be represented also in areas belonging to the dorsal stream (Sakata *et al.*, 1995; Murata *et al.*, 1996). This has lead to a rethinking of the two visual streams in terms of their purpose: the ventral stream is concerned with the conscious perception of objects (and their

---

[2] 'Detecting' is used here in the meaning that a neurone fires, when the relevant feature appears in the world.

position), while the dorsal stream processes the same attributes (shape, position and motion), but for a different purpose: to guide motor actions without conscious awareness (Goodale and Milner, 1992; Milner and Goodale, 1993). Indeed the small number of connections between the two streams (Felleman and Van Essen, 1991; Jouve et al., 1998; Young, 1992), and the lack of conscious shape awareness of patients with lesions in the ventral stream (Milner et al., 1991; Milner, 1997) supports the idea of the ventral stream being the prime site of *conscious* shape perception.

1.2. Investigations of higher visual functions

This distributed hierarchical 'text book' model of visual processing is thus still undergoing substantial changes. What is important about it, is that despite its obvious limitations, it sketches a basic image of the visual system that allow us to start thinking about the neural basis of 'higher', psychological phenomena. It creates the necessary, minimal foundations for going beyond mere neurophysiology to what has been termed 'cognitive neuroscience': the investigation of cognitive functions using neuroscientific methods. Indeed, in the last two decades, an increasing number of investigators have used the basic knowledge of the neural architecture of the visual system to study successfully 'higher' aspects of vision. For instance, the neural correlate[3] of visual working memory (Brown and Xiang, 1998; Fuster 1981), of visual long term memory and its recall (e.g. Naya et al., 1996), of visual attention (Duncan, 1998; Desimone, 1998), of binocular rivalry (Logothetis, 1998) and of visual masking (Macknik and Livingstone, 1998; Kovács et al., 1995; Rolls et al., 1999) have been investigated with great success.

Visual perception refers to the process in which a visual stimulus can create a conscious percept. To investigate the neural correlates of perception, it is necessary to understand (i) how the brain processes a particular visual stimulus and (ii) to understand which part of this neural processing is statistically associated with the report of a particular percept. The present thesis will investigate both these points for a particular perceptual paradigm.

Two lines of investigations have so far been particularly prominent in their effort to understand the neural correlates of visual perception. One investigates the

---

[3] 'neural correlate' refers to observable neural activity which occurs specifically, when a certain mental function is achieved, and which is *thought* to underlie this mental activity.

neural correlates of motion perception (e.g. Shalden and Newsome, 1996; Celebrini and Newsome, 1995; Shalden et al., 1996; Britten et al., 1996). The other investigates visual perception during binocular rivalry (Logothetis, 1998, reviewed in Chapter 2). Both of these attempts have compared neural activity with the report of a particular percept. Indeed, since the percept itself cannot be observed by anyone except the subject that perceives it, perception itself is beyond scientific enquiry. Yet, subjects can be asked to report their perception, and scientific investigation can be successful at defining the neural correlates of the *report* of a percept: a strongly related question. The present thesis will approach the question of the neural correlates of visual perception by in effect investigating the neural correlates of perceptual report.

1.3. The present thesis: examining the neural correlates of visual perception in RSVP

The originality of the present thesis will be to be the first investigation of the neural correlates of visual perception using a paradigm that has not been used in that context before: Rapid Serial Visual Presentation (RSVP). RSVP is a mode of stimulus presentation in which images are presented as an accelerated slide show: individual images follow and replace each other on a screen at a rapid rate of up to 72 images per second. Unlike the images of a movie, which are also presented one after another in a continuous sequence, stimuli in RSVP are not successive snapshots of a continuous sequence of events. Instead, essentially unrelated images are presented one after another. What makes RSVP suitable for the investigation of visual perception is that varying the rate at which images are presented affects how much perception of individual images occurs. At slow rates, images are perceived clearly, while at very rapid rates, individual images are almost impossible to perceive.

Unlike in functional imaging studies, which at present lack the possibility to investigate the function of single neurones, single cell recordings necessitate the choice of a brain area in which to search for neural correlates of visual perception. The brain area STSa is amongst the very last stages of the ventral stream. It contains single cells that respond selectively to complex patterns such as faces and the activity of single neurones in STSa correlates with perception during binocular rivalry (Sheinberg and Logothetis, 1997; Logothetis, 1998). In addition, single neurones in STSa correlate with the perceptual effects of visual masking (Rolls et al., 1999). It

may therefore be reasonably hypothesised, that STSa is involved in the perception of images, and is therefore a promising anatomical starting point for the investigation of neural correlates of visual perception in RSVP sequences. It is hence in STSa that the single cell recordings of this thesis will be performed.

The hypothesis that single neurone responses in STSa may be involved in the perception of images will be tested using the following approach. First, the effect of changing the presentation rate of an RSVP sequence on the neural activity of single STSa neurones will be measured. Second, by asking human subjects to report individual images in RSVP sequences, the effect of changing the presentation rate of an RSVP sequence on perception will be measured. By comparing the two, one can test, if the changes in STSa single cell activity correlate with changes in perception. In a second step, inter-stimulus gaps will be introduced between images in RSVP sequences and the effect of these gaps on single cell activity in STSa and perception will be measured and compared, to clarify further the role that STSa may play in perception.

1.4. The organisation of this thesis

Chapter 2 and 3 will review the literature on binocular rivalry and visual masking respectively. Binocular rivalry will be reviewed, because it is an excellent example of an investigation of the neural correlates of visual perception. Visual masking will be reviewed because of its similarities to RSVP. Chapter 4 will attempt to extract a new conceptual framework from these two reviews: stimulus competition in the visual system. Chapter 5 will introduce RSVP, and present some of the psychological findings of RSVP. Chapter 6 will present the general physiological methods used in the thesis to investigate the effect of RSVP on single cell activity in STSa. Chapter 7 will be the first experimental chapter of this thesis, and will illustrate the first findings obtained by measuring the effect of RSVP presentation rate on single cell activity in STSa. Chapter 8 will use information theory to analyse the results of Chapter 7. In Chapter 9, the effects of RSVP presentation rate on human perception will be measured, and in Chapter 10, the effects of RSVP presentation rate on human perception and single cells in RSVP will be compared, to conclude the investigation of the effects of RSVP presentation rate. Chapter 11 will illustrate how introducing interstimulus gaps affects the responses of STSa neurones during the presentation of

RSVP sequences. Chapter 12 will demonstrate how interstimulus gaps affect different aspects of the perception of RSVP sequences differently, and will show that the perception of image identity but not other aspects of perception correlate with the activity of single neurones in STSa. Chapter 13 will then conclude this thesis by presenting an overview of the main findings.

Parts of this thesis are currently in press (Keysers et al., 2000). This publication incorporates the main findings of chapter 7, 9 and 10. In addition, a number of other publications are in preparation. We will attempt to adapt Chapter 8 to a publication on information theory, Chapter 11 and 12 to a publication on visual persistence. Finally, *Nature Neuroscience* has commissioned David Perrett and me to write a review article on stimulus competition in the temporal domain: Chapter 2-5 will provide the basis for this review paper.

The findings of this thesis have also been presented at international conferences, namely the ENA meeting in Berlin (1998), the ASSC3 meeting in London (Ontario, 1999) and the IBRO meeting in Jerusalem (1999). This year, the findings of Chapter 11 and 12 will also be presented at the Society for Neuroscience meeting in New Orleans (2000) and a Fondation des Treuilles meeting in Tourtour (France, 2000).

2. Stimulus competition in the visual system: I. Binocular Rivalry

There are a number of instances in the visual system in which stimuli compete against each other: binocular rivalry, masking and attention. In all three cases, different stimuli compete against each other to be represented in the visual system and be perceived consciously. These instances are usually seen as quite distinct phenomena, the goal of the following few chapters is to review these phenomena separately, and then to show how they are all examples of a common phenomenon of stimulus competition in the visual system. This general phenomenon will then later be used to explain the results of the empirical investigations of this thesis. Indeed, when the word 'stimulus competition' will be used, it will be used as a short-hand for 'competition between the neural interpretations of the visual input'.

2.1. Introduction to binocular rivalry

In binocular rivalry, two different stimuli are presented to the two eyes: a situation called 'dichoptic' stimulation. The two stimulus themselves remain constant over time, but perception alternates stochastically between different states. Most of the time, either one or the other stimulus is perceived, while the other one is fully suppressed. Under normal conditions the two percepts between which the system alternates are the two dichoptically presented stimuli. In less usual cases, the precepts can be the result of a regrouping of features from the two eyes, as will be seen later. Between such exclusive phases, there are short intervals of "piecemeal" perception, in which a bit of both stimuli can be perceived. Binocular rivalry has been extensively studied from a psychophysical point of view, and in recent years the brain mechanisms underlying the phenomenon have been investigated using single cell recordings in the macaque and fMRI and EMG in human subjects. A central question of the present thesis will be to investigate the relationship between single cells in the anterior superior temporal sulcus (STSa) and the perception of images. Binocular rivalry is a powerful tool in the study of these neural correlate of consciousness

because it dissociates the physical stimuli on one hand, and the subjective percept on the other: the stimuli are presented constantly while perception, for no apparent external reason, fluctuate endogenously between two alternative percepts. A given type of brain activity can thus only participate directly in perception if it follows these changes in perception. The paradigm used to investigate brain activity during rivalry is essentially the same in most experiments. An indicator of brain activity is found that is specific for one of the two monocular stimuli (e.g. a single cell selectively responding to one of the patterns or a given set of voxels in an fMRI). The indicator is then quantified under monocular vision of the two stimuli (A) and (B) separately and to their joint, dichoptic, binocular rivalling presentation (AB). If the part of the brain the indicator is taken from is correlated with perception, when the subject reports seeing only the stimulus A, the brain activity should correspond to that in the condition (A), if he reports seeing B, it should resemble that in condition (B). If on the other side, the brain activity is only reflecting the stimuli, not the percept, the activity should remain constant, whatever the perceptual report.

For the present thesis, binocular rivalry is of double interest. First because it studies neural correlates of visual perception and secondly because it has something in common with RSVP: in both cases the processing of one stimulus suffers from the presentation of other stimuli.

2.2. Psychophysical literature

2.2.1. Two stimuli compete, not two eyes

Originally it was thought that binocular rivalry is a very specific effect that depends on the competition of the two ocular channels. The idea was that one eye or the other is refused access to higher visual processing. This idea has been proven wrong. I will here review a number of studies that illustrate why this belief was is wrong.

Diaz-Caneja, (1928) showed, that the stable unified percepts emerging from binocular rivalry need not be either the left or the right visual input. They can be a 'meaningful' combination of parts of both. The effect is illustrated in Fig. 2.1a. If binocular rivalry switched one eye on and off, the percept should represent one of the two ocular inputs. Instead, often, either a bilateral sun-ray or bilateral horizontal stripe

pattern is perceived. The gestalt laws can therefore sometimes override the eye of origin, and group elements from different eyes.

This interocular grouping in binocular rivalry was confirmed much later by Kovács et al. (1996). They measured the proportion of time a unified pattern was perceived during binocular rivalry. Fig. 2.1b illustrates the kind of stimuli their used. The 'unified' percept was either a matrix of all red or all yellow dots. That percept arose both if one eye saw an all red matrix and one an all yellow matrix (upper) and when neither eye receives a unified stimulus, but interocular grouping was necessary. Interestingly, in the latter case, a unified percept arose 47% of the time, which is much more often than expected by chance if one would postulate only a local competition between the two dot-colours at every dot location, for instance in V1, where receptive fields are small. In addition, 47% is rather close to the 60% unified perception arising when the stimuli are themselves unified. It therefore appears, that the competition between the two stimuli occurs not in early visual areas (V1/V2) alone, since the precepts accord to higher levels of classification (such as sunray patterns, etc) which may not be explicitly represented in V1/V2.

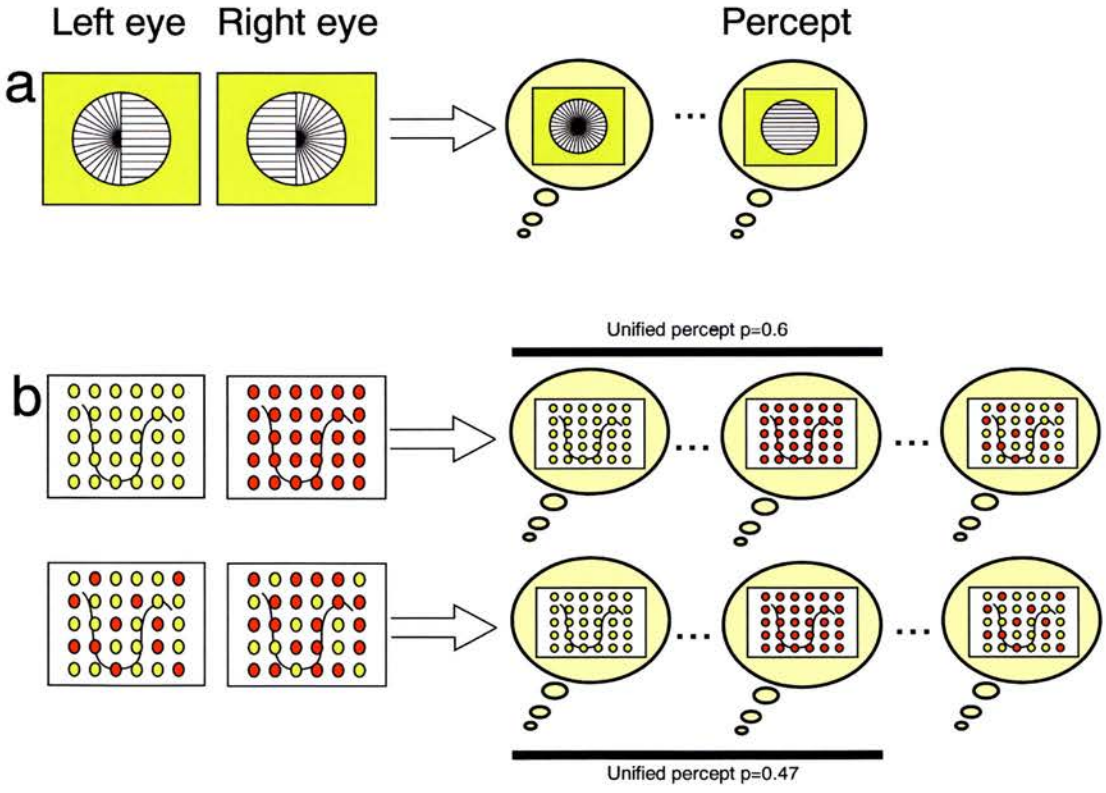


Figure 2.1: (a) Stimuli of the type used by Diaz-Caneja (1928). If fused binocularly, a bilateral radial sun-ray or a bilateral horizontal grating pattern is often perceived. This is incompatible with the idea of competition between the eyes. (It can be experienced to some extend by divergent fusion of the two images, but minimal movement will make the two grating move and brake their unity). (b) Stimuli used by Kovács et al (1996). The proportion of unified percept is surprisingly high in the lower condition ($p=0.47$) and is certainly higher than expected by a chance pairwise piecemeal combination of the two rivaling patterns. This illustrates, that while eye of origin is an important criterion for selecting what is going to be perceived, it is not the only one: gestalt laws too, although to a somewhat lesser extent ($0.47<0.6$) can determine the result of the rivalry.

The idea that eye of origin plays a limited role (60% compared with 47% of prevalence) is also confirmed by Logothetis et al. (1996). They presented orthogonal gratings to the two eyes of human subjects, and measured the distribution of the duration of perception of either one of two gratings. Then they repeated the same experiment, but every 300ms, they swapped the presentation side of the two gratings. If eye of origin was the criterion of selection, then this should have resulted in a dramatic change of the percept duration distribution. Yet, the percept duration was unchanged, indicating that in binocular rivalry a stimulus is selected for perception and not an eye.

Andrews and Purves (1997) quite convincingly show that competition between neural interpretations of the stimuli and not eye-competition is the main principle underlying the phenomena of binocular rivalry by showing that using appropriate stimuli (e.g. two orthogonal gratings, one red, one blue on a gray background, all isoluminent), perception will alternate between the two monocularly rivalling stimuli, with a pattern of fluctuation similar to that of binocular rivalry, although the absolute frequency of switches between seeing only the red or blue lines was less than in binocular rivalry. Further they demonstrated that binocular rivalry and monocular rivalry can interact with each other: they presented their monocular rivalling pattern (i.e. both red and blue lines) in an annulus surrounding a centre in which the same components were presented in a binocular rivalling fashion. Not only did they find percept alternation in both the binocular and monocular rivalling sections, but the switches in percept occur simultaneously in both sections, at a frequency intermediate between that measured for monocular and binocular rivalry alone. This indicates that the mechanisms underlying binocular rivalry are not an

isolated instance of competition between two eyes, but are interacting with other mechanisms of stimulus competition, such as those observed in monocular rivalry.

2.2.2. Rivalry takes about 150ms

Wolfe (1984) investigated the temporal dynamics of binocular rivalry. If two orthogonal gratings are flashed simultaneously two the two eyes, they are perceived as merged (i.e. piecemeal rivalry or "checkerboard") if flashed for less than 150ms. Only with stimuli lasting longer than 150ms do the two stimuli engage in rivalry.

In flash suppression, one stimulus is given to one eye, while at first, no stimulus is presented to the second eye. After a while, a new stimulus is presented to the second eye, while the first stimulus continuous to be presented to the first eye. If the first flash had been presented for more than 150ms, the new stimulus will always be perceived as soon as it is flashed. If the first stimulus was presented less than 150ms, piecemeal rivalry will occur. It therefore appears, that rivalry takes 150ms before the competition between the stimuli results in suppression.

2.2.3. Shape and movement have independent access to perception under rivalry

Andrews and Blakemore (1999) using binocular rivalry could show that movement and object perception are differently affected by binocular rivalry. They presented two perpendicular diagonal gratings, moving such that, if presented alone through an aperture a component motion perpendicular to the grating was perceived (Fig. 2.2a), while if presented together to one eye, a pattern motion created by the resulting plaid, and equal to the common vector component of the two component motion was perceived. If two different gratings are presented to the two eyes, rivalry occurs, and only one of the gratings is perceived in most cases. On those trials with complete rivalry, one would expect, that if only one grating is perceived, the motion of that grating should be equal to the monocular component motion of that pattern. But not so: in 50% of the cases, although only one grating was perceived, the perceived motion was that normally associated with both patterns seen simultaneously in one eye. Hence, while the pattern perception of one of the patterns was suppressed, the motion perception of that same pattern was not, indicating an independent access to perception for the motion and pattern recognition system.

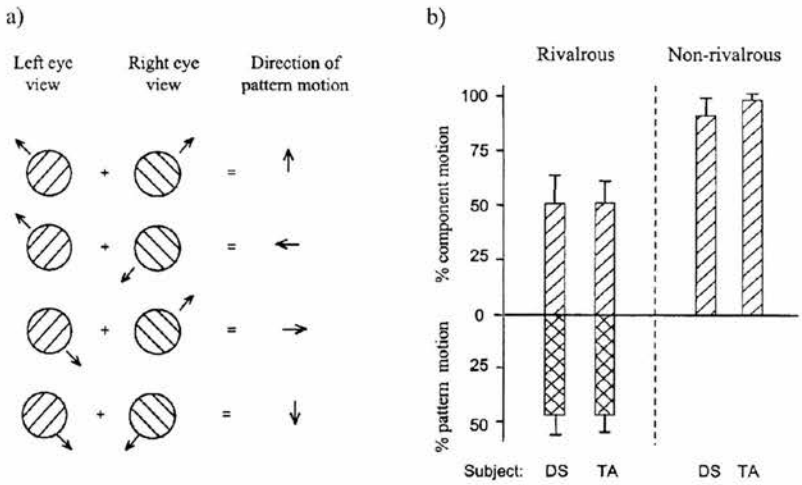


Figure 2.2: (a) Diagonal drifting gratings presented to the two eyes in the study by Andrews and Blakemore (1999), together with the direction of motion that is perceived if the stimulus presented to the left or right eye was seen alone (i.e. component motion, as indicated by the black arrow next to the pattern), and the motion that would be perceived if the two stimuli were seen together, forming a plaid (rightmost arrows). (b) Perceptual report in conditions of rivalrous presentation (left) as described in (a), or nonrivalrous conditions, where the same drifting grating is presented to the two eyes. As can be seen, in the non-rivalrous conditions, only component motion is perceived, as is generally the case for such "aperture" gratings. In the rivalrous conditions though, in about 50% of the cases, the pattern motion is perceived, as would be the case, if in these 50% of cases the two gratings were seen at the same time, forming a plaid. Yet, all trials in which the observers reported seeing any combination of the two gratings were excluded from the analysis. Hence it is likely, that although the pattern recognition system perceptually suppressed one of the gratings, the motion system still had access to that second grating, and has an indipendent access to perception. [Adapted from Andrews and Blakemore, 1999].

2.2.4. Monkeys and human show similar patterns of rivalry

Before measuring neural correlates of binocular rivalry in the monkey, the group around Logothetis showed that binocular rivalry is almost indistinguishable in humans and monkeys. Using a variety of different stimuli, they showed that the normalised distribution of the durations between changes in percept (i.e. the dominance time) is identical in humans and monkeys (Leopold and Logothetis 1996, Sheinberg and Logothetis 1997) and that the effect of spatial frequency and contrast on dominance time is identical (Sheinberg and Logothetis, 1997; Logothetis 1998). Flash-suppression is a variant of binocular rivalry. A pattern is presented continuously to a first eye, while nothing is presented to the second eye for at least 150ms. Suddenly, a second stimulus is presented to the second eye, while the first eye continues to be presented with the first pattern. For humans, perception switches immediately to the

second pattern, as soon as it is presented. For monkeys, trained to report their percept, the result is identical (Sheinberg and Logothetis 1997). These remarkable similarities underline the similarities of the visual systems in the two species, and legitimates the comparison of macaque monkeys and humans.

2.3. Single neurone investigations

2.3.1. Early visual cortex (V1/V2)

Monkeys were presented binocularly with orthogonal gratings and pressed a left lever if they perceived a grating tilted to the left, a right lever, if they perceived a right tilted grating, and refrained from pressing any lever if they perceived a patchwork of right and left tilted gratings. Only 6/33 neurones in V1/V2 showed changes in firing rate correlated with the monkey's perceptual report (Leopold and Logothetis, 1996). Changes in firing rate during percept switch in binocular rivalry are thus rare in V1/V2.

In strabismic cats, the synchrony of pairs of neurones representing the same pattern in Area 17/18 – synchrony being quantified as the cross-correlation of firing rate between pairs of neurones - follows strongly the percept (Fries et al., 1997). In those strabismic cats, one eye becomes dominant, and if two gratings of identical contrast moving in opposite directions are presented to the two eyes, the dominant eye will determine the direction of the optokinetic nystagmus and presumed perception in over 90% of the trials. Recording multiunit activity in area 17/18, about 30% of the pairs of multiunit recording cites showed synchronous activity if stimulated monocularly with a grating of preferred orientation. If a pair of multiunits responded to stimuli presented in the dominant eye, the central peak of the cross-correlogram between the firing rate of the two multi-units was enhanced by on average 55% when an orthogonal grating was presented to the suppressed eye. If a synchronized pair of multiunits was responding to a grating in the subdominant eye, their synchrony decreased by 31% upon presentation of an orthogonal grating in the dominant eye. Of the 55 pairs that had shown no synchronisation upon monocular presentation, 11 became significantly synchronized during rivalry. Neurones contributing to perception therefore seem to be tagged by synchronising their activity. In contrast to these clear

effects on synchrony, the effect on firing rate were small and surprising: both the firing rate of multiunit recording cites responsive to the dominant and the subdominant eye were *reduced* if a orthogonal grating was presented to the contralateral eye (by 9.2%, and 3.2% respectively). It is difficult to conceive how a reduction of activity in neurones representing the dominant stimulus can participate in making it dominant.

It is tempting therefore to believe, that the neurones representing the component parts of the dominant stimulus, by becoming synchronized, will have an advantage in recruiting neurones in the subsequent stages of processing. Nevertheless, the data in strabismic cats is unfortunately hard to generalise to normal subjects. In these animals, brain cells almost never experience congruent binocular information, which might result in a specific anti-hebbian training and lead to specific mechanisms to deal with the permanent incongruity of the binocular input. In normal animals the two eyes usually provide congruent information, which will result in a very different neuronal training. Nevertheless, MEG data that will be presented later supports the idea that synchronisation in early human visual cortex may correlate with perception during binocular rivalry.

2.3.2. Higher visual cortex:

V4 for shape processing and MT for movement processing seem to be the brain areas in which the two rivalling neural interpretations of the stimuli compete most against each other. In V4 38% (Leopold and Logothetis, 1996) of the neurones respond differently when the monkey reported the preferred or the non-preferred stimulus for the cell. ~25% of cells fired more when their preferred stimulus was reported, and ~13% fired more when their *non-preferred* stimulus was reported. These ~13% are likely to be involved in suppressing the perception of their preferred stimulus interpretation, and indicate that V4 is probably the cortical stage at which the active suppression of a static stimulus occurs. At this stage neurones are binocular, so that stimulus competition seems to occur after the visual channels have fused. In MT the situation is similar, with ~20% responding more to the perceived and ~20% more to the suppressed moving grating (Logothetis and Schall, 1989). In IT and the anterior parts of the superior temporal sulcus (STSa), it appears that 52/58, i.e. 90% fire more to their preferred stimulus interpretation if it is perceived than if it is suppressed, and

none did the reverse. While the responses are smaller, when the stimulus is not perceived, the responses are nevertheless not quite null, and in 10% of the cells, the response does not change with perception. It remains to be understood, why the remaining IT activity representing the suppressed stimulus does not create perception. Also, in flash suppression, if the cell's non-preferred stimulus is shown to one eye for a while, and the cells' preferred stimulus is then flashed to the second eye, the cells' latency to respond to the preferred stimulus is ~60ms later than the cells' normal latency, indicating that it takes ~60ms for the winner-take all competition to switch off the old stimulus and turn on the new stimulus (Sheinberg and Logothetis, 1997, confirmed by personal communication, see Fig. 2.3).

2.3.3. Single cell conclusions

The single cell data indicates that early in the visual system (V1/V2) both of the 2 rivalling interpretations of the dichoptic visual input are represented, the perceived interpretation being possibly tagged by the synchronisation of its ensembles. In intermediate levels of processing for shape and motion (V4/MT), an active suppression takes place, and at the level of IT and STSa, a strong correlation of firing rate with perception is observed. This evidence is compatible with the idea that perception may arise from activity in IT or STSa or later in the system, based on the output of IT or STSs. Alternatively, perception could be based on a minority of cells throughout the visual system, namely within the ~20% of V1/V2 cells that follow the percept, the ~20% in MT or V4 etc. The single cell data also indicates, that mechanisms in the early visual system exist that allow 2 stimulus interpretations to be represented at the same time. Later in the visual system a strong competition between the two equally likely interpretations can be observed, leading in a winner-take-all manner, to the exclusive representation of one of the two interpretations at the expenses of the other.

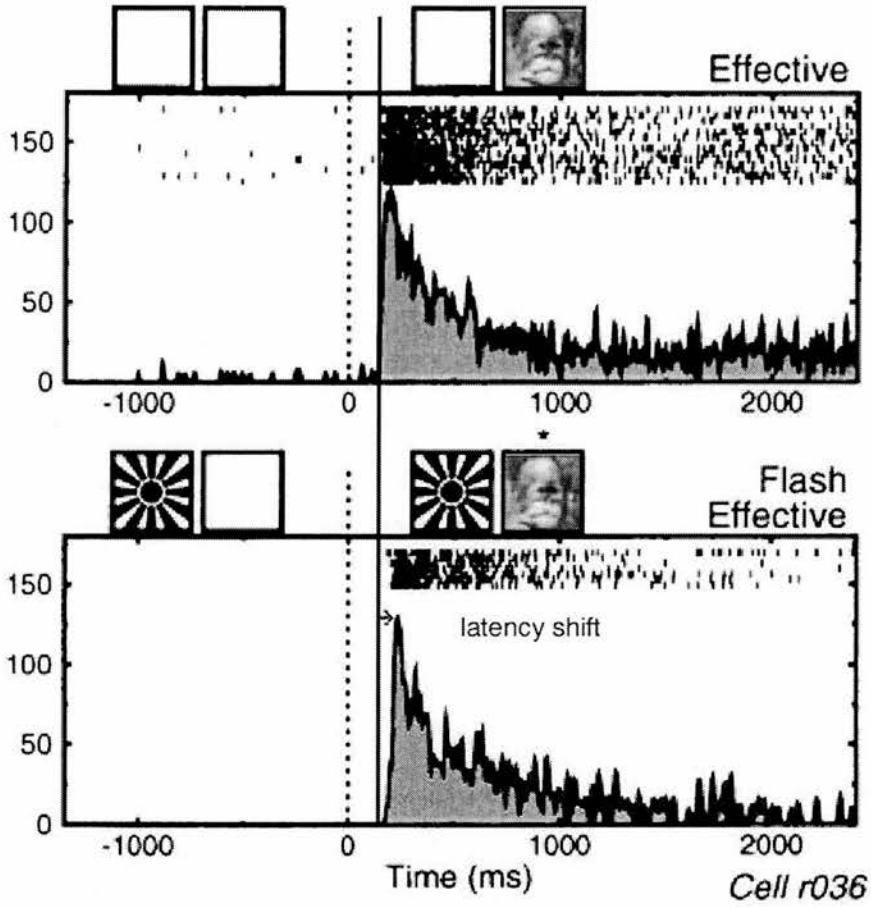


Figure 2.3: Latency shift in flash suppression. Top: response to the effective stimulus being presented monocular and for a single cell. The black vertical line represents the latency of the cell to respond to an effective stimulus. Bottom: response of the same cell under flash suppression. First, the sunray pattern is presented to the right eye while nothing is presented to the left eye. Then, at t=0, the cells preferred stimulus is introduced to the left eye, while the right eye continuous to see the sunray pattern. The monkey reports perceiving the monkey face (as denoted by the star above the stimulus), and the firing resembles that under monocular viewing, except for a shift in latency by ~60ms [Adapted from Sheinberg and Logothetis, 1997. Sheinberg confirmed to me, that this latency shift was observed in all neurones].

2.4. Functional imaging studies (fMRI + MEG).

fMRI studies confirm the fact that in the later stages of the visual cortex one of the two stimulus interpretations is almost completely suppresed. Tong et al. (1998) have identified two brain areas that respond to the sight of particular objects. The fusiform face area (FFA) is activated following the presentation of faces and the parahippocampal place area (PPA) is activated after the presentation of places such as a house. If a house and a face are presented under rivalry conditions, while the subject signals seeing a face (i.e. one of the two rivalling interpretations), the activity of the

FFA is as high as if a face was presented alone, while the PPA shows no activation. When the subject reports seeing a place, his PPA is as active as if the place was shown alone, and the FFA is inactive. Lumer and Rees (1999) presented a face to one eye, and a moving grating to the other eye, and showed that the activity in the fusiform gyrus and the superior parietal cortex (involved in movement processing) are related to perception of motion and face, while earlier areas (V1/V2) are not. In addition, they showed that a number of other, non-visual areas in the frontal and prefrontal cortex are activated by the rivalry in a way correlated to perception. They interpret this to indicate that an interaction between these areas and the more strictly visual areas may be important for the emergence of perception. Since in their experiment subjects did not have to report their perception, this frontal activity cannot be attributed to explicit decision make or motor response initiation. Nevertheless, it may reflect neuronal computations occurring after perception, which receive their input from the visual areas mediating perception.

The percept-modulated activity found in frontal and prefrontal areas, beyond the classically visual areas is confirmed in magnetic encephalography (MEG) studies (Tononi et al. 1998). In an elegant design, they used two gratings flashing at different frequencies (f1 and f2) to the two eyes. The neural activity created by the two stimuli can be separately measured at any of the 148 MEG recording channels as a peak at frequency f1 or f2 in the Fourier transform of the signal at that location. The activity was distributed according to a horseshoe pattern, encompassing occipital, parietal, temporal and frontal areas. The power at the frequency of the stimulus that the subject reported not seeing was lower when compared with its power when it was perceived, in accordance with the single cell findings. Interestingly, the difference in power was only 50-85% of the signal magnitude obtained if the stimulus was presented alone, indicating that while binocular rivalry results in a strong modulation of the signal, it does not completely abolish neural representations of the suppressed stimulus even in the more anterior areas of the brain. In addition, in accordance with the experiments in strabismic cats, the signal in occipital locations was to some extent modulated by perception. Since MEG measures synchronized neuronal activity, this is compatible with the idea that the amount of synchronization is increased in occipital regions when a stimulus is perceived.

2.5. Overall conclusions

The most surprising finding in binocular rivalry is the fact that not the eyes nor the two stimuli compete directly against each other, but rather two equally likely neural interpretations of the dichoptic visual input. These two interpretations usually are the two stimuli, but sometimes, the rivalling interpretations each incorporate elements of both stimuli. This is shown in numerous psychophysical experiments that show that perception will tend to alternate between two meaningful stimulus interpretations even if the meaningful interpretations have to be created by the combination of features presented separately to the two eyes. The same finding is supported by the physiology. Early in the visual cortex, both the perceived and the suppressed ocular input are represented. It is only in later visual areas that the features will compete against each other, and a single "winner-take all" interpretation emerges in 90% of the cells. The eye of origin only seems to play a minor role in deciding which features will win, and which features will be bound with which features, as demonstrated by the study of Kovács et al. (1998). Hence the powerful competitive mechanisms unravelled by binocular rivalry are likely to be much more general. As demonstrated by Andrews and Purves (1997), when incompatible stimuli arrive through the same eye, they will be subject to a perceptual competition similar and linked to that occurring if they arrive through different eyes. It is likely that the neural competition revealed by the investigation of binocular rivalry also applies to such cases of monocular rivalry, and even to figure ground segregation such as the vase-face figure. The value of the experiments on binocular rivalry is then to make us aware of the strong competitive mechanisms that are build into the brain, mechanisms that are valid beyond binocular rivalry.

A further interesting aspect of the investigations of binocular rivalry are the implications for the mind-brain question. Perception in binocular rivalry correlates with the brain activity higher in the visual system, including inferior temporal, some frontal and prefrontal areas. Perception correlates much less with the activity earlier in the visual cortex, although about 20% of the cells in these areas would be adequate for supporting perception. For synchrony it appears that in early visual areas synchrony, independently of firing rate, correlates with conscious perception. The relationship between synchrony and perception is unclear. Synchronised neuronal activity will be more effective at driving postsynaptic neurones due to temporal summation of

synaptic potentials. Perception may thus either be dependent upon synchronization *per se*, or to increased firing rate in the neurones downstream. In this second perspective, synchronization may be an economic way to bias the competition between the neurone populations representing the two stimulus interpretations since it does not require an energetically more expensive increase of firing rate and allows to maintain the representation of the non-perceived interpretation in terms of firing rate, while denying its access to the following cortical levels. The experiments by Andrews and Blakemore (1999) also teach us, that movement and shape may reach perception separately, with shape perception more affected by suppression than movement perception.

3. Stimulus competition in the visual system: II. Persistence and Masking

In binocular rivalry two stimuli are presented *simultaneously* to the observer, and two neural interpretations of these stimuli compete for representation in consciousness and higher brain areas. In masking two stimuli are presented one after the other, *not overlapping in time*. Masking conditions thus appears to be fundamentally different from binocular rivalry. Yet 'persistence' links the two phenomena together. Persistence refers to the fact, that unless masking occurs, a stimulus will continue to be represented both in consciousness and in the brain for a certain amount of time after it has been physically removed. This persistence of a stimulus is still there, when the second stimulus is presented in the masking paradigm. Seen in this way, in the brain, in masking – as in binocular rivalry – the neural **representation** of two *simultaneously* **re**presented stimuli compete against each other. This Chapter will first review persistence, and then masking. Both will be reviewed in some detail, because they are extremely relevant to the experimental findings reported in this thesis.

3.1. Persistence

Persistence refers to the fact that a stimulus remains visible to the observer for some time after it has been physically removed from the retina. Coltheart (1980a,b) in two excellent reviews points out with remarkable clarity, that three types of persistence need to be distinguished. "There is visible persistence: visual stimuli continue to be visible for some time after their physical offset. There is neural persistence: at various stages in the visual system, neural activity evoked by a stimulus continues to occur for some time after stimulus offset. There is informational persistence ('iconic memory'): the sensory information contained in a visual display remains available to an observer for some time after stimulus offset" (Coltheart, 1980a p67). The relationship between these three types of persistence, in particular between information and visible persistence, remains unclear and heavily debated (e.g. Loftus and Irwin, 1998;

Coltheart, 1980a,b). Within the present context, going into the details of that debate would be confusing. What is important here, is to show that persistence is an important, well established phenomenon in vision, that can be observed readily using a broad spectrum of paradigm. The differences between the findings are less important as the general undisputed finding that as far as the neural system or an observer is concerned, a stimulus doesn't disappear when it disappears. The representation of the stimulus remains present in the mind's eye for a certain duration beyond the stimulus' physical termination. The estimates of the duration of persistence vary largely, from almost zero to about 1s depending on the paradigm. Details of the stimulus, such as luminance, duration, spatial frequency influence the duration of persistence, and influence different estimates in different ways. Presenting an adequate pattern (the mask) after the stimulus can terminate this persistence. While all these differences are very interesting, the single most important finding is, that unless a mask is presented after a brief stimulus, just turning off the stimulus on the screen will not turn off the stimulus in the brain or in the mind. Some of the findings will be reviewed separately for the 3 types of persistence.

3.1.1. Visible persistence

Visible persistence is measured through a number of paradigms. What is common to all these paradigms, is that they try to measure the duration of the *percept* of a stimulus. They therefore rely not on a psychophysical response of the subject that only reflect how much information the subject has about the stimulus, but rather on asking the subject to respond to what the percept "looks" like. Seven different paradigms have been used to measure the duration of a percept.

(1) Onset-offset reaction time. Subjects are presented with a brief stimulus, e.g. a grating. In a first set of trials, the subject is asked to press a button as soon as they perceive the start of the stimulus. In a second set of trials, the subject has to press a button when he perceives the end of the stimulus. Baro et al. (1992) measured the perceived duration of a vertical grating using this paradigm, and found that the interval between on-RT and off-RT is longer than stimulus duration by only 20-40ms. Like all other measures used, this task critically depends on the criterion used by the observer for offset. Subjectively, the stimulus seems to fade of progressively and

exponentially after its physical offset. When does the subject press the button? As soon as he notices a decay, or when the image has decayed to zero? The criterion used will determine the offset RT and the persistence. Persistences as short as 20-40ms in Baro et al. (1992) may be due to very high criteria. Divergences between different papers, often using just one or two subjects, can thus depend on differences in criteria. Baro et al. (1992) also found, that although RTs do change as a function of grating frequency and duration, both the onset and offset-RT vary by the same amount, resulting in an unchanged perceived duration. These findings have to be taken with caution, because for humans it is easy to press a button to the onset of a brief stimulus, but it is hard to suppress the tendency to press to the beginning, and press to the end in stead. Hence, off-RT will be increased for short stimuli (see Di Lollo et al., 2000). The phosphor decay times of the presentation screen may also seemingly increase persistence – a point that applies to all measures of persistence.

(2) judgment of synchrony or the interclick interval. Here, the subjects are asked to synchronise a brief event (an auditory click or a visual flash) with the perceived onset and offset of a stimulus. Due to differences in processing speed in the auditory and visual modality, an auditory click is generally perceived about 30ms faster then a visual stimulus (Haber and Standing, 1970), but since the same latency difference seems to apply to the onset and offset of the stimulus, the inter-click time gives a valuable estimate of perceived visual duration. Using this method, Haber and Standing (1970) found persistence to range from 0ms for a 700ms flash presented after a relatively bright adaptation field to 400ms for a 20ms flash of same intensity presented without adaptation field (Fig. 3.1). They also investigated the effect of presenting a pattern (composed of small letters) or a flash of light after the stimulus, and found that presenting the pattern, but not the flash of light resulted in inter-click intervals identical to the stimulus onset asynchrony (SOA) between stimulus and pattern. This indicates that the pattern but not the flash could terminate the persistence of the stimulus. Bowen et al. (1974) asked subjects to align a brief visual flash presented 2deg away from the stimulus. They found persistence (i.e. the time interval between physical and reported end of stimulus) in the order of 100 to 350ms, decreasing with both stimulus duration and intensity.

Together, these two variants of the synchrony method yields persistence times in the hundreds of ms, and shows two conterintuitive findings typical for visual

persistence (as opposed to information persistence): the inverse duration and inverse intensity effect. The inverse intensity effect states, that as the luminance of a stimulus or a background *increases*, the duration of visible persistence *decreases*. The second effect is the inverse duration effect: the *longer* a stimulus, the *shorter* its persistence. This effect generally holds for stimulus durations up to 100 to 200ms, thereafter persistence has reached a minimum (see Fig. 3.1). If very bright stimuli are used, these relations do not apply anymore, probably due to the fact, that positive afterimages are perceived rather than visible persistence proper (Long and Sakitt, 1981).

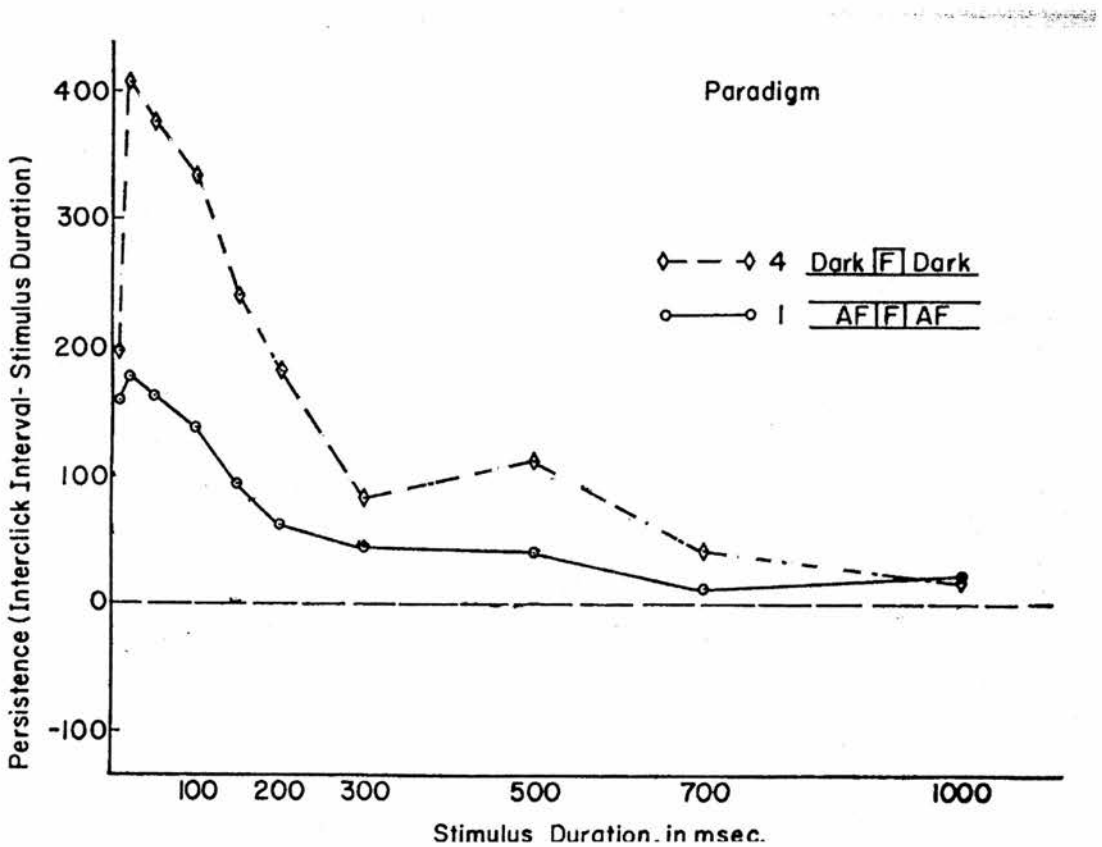


Figure 3.1: Visible persistence as measured by the synchronisation of a click on the perceived onset and offset of a stimulus flash (F) as a function of stimulus duration and adaptation field. Both before and after the flash, either a dark or a lightly illuminated adaptation field (AF) was presented. Of the 4 conditions tested in the experiment, 2 are particularly interesting and shown here. In condition 1, an adaptation field consisting of a circle of modest luminance centred on the fixation point is presented both before and after the stimulus, which was a 3x3 matrix of numbers. In condition 4, no adaptation field (=Dark) was presented, and the screen was dark except for the stimulus. As can be seen, in all cases, the visual persistence (y-axis) is higher than 0, indicating visual persistence. The duration of the persistence is inversely correlated with background illumination (inverse intensity effect) and also inversely correlated with stimulus duration (inverse duration effect) except for the

shortest stimulus (<10ms), for which the energy might be too low yet to produce a proper perceptual experience [Adapted from Haber and Standing, 1970].

(3) phenomenal continuity. A single stimulus is turned on and off over and over again. The shortest gap between two consecutive occurrences of the stimulus, for which an observer perceives that the stimulus completely disappeared between two presentations is then taken as the duration of persistence. In this paradigm, with increasing gap, subjects first perceive flicker, but despite this fluctuation in brightness, the stimulus seems to remain on the screen during the entire time. Then with even longer gaps, it becomes clearly perceivable that the stimulus pattern actually does vanish from the screen. Subjects are instructed to ignore the fluctuations in brightness (=flicker), and just to focus on whether the stimulus completely disappears between two cycles. This method applied for instance by Bowling and Lovegrove (1980) leads to estimates of persistence in the order of 150ms (for a 300ms grating of 1c/deg) to 300ms (for a 50ms grating of 12c/deg), with high frequency gratings persisting longer.

(4) temporal integration of form parts. A stimulus is broken into parts, the parts are shown at different times, and the longest inter-stimulus interval at which the parts fuse to a single stimulus is taken as the duration of persistence. Depending on the complexity of the stimulus, the estimates of persistence duration vary (e.g. Di Lollo 1980).

(5) Stroboscopic illumination of a moving object. A rotating object is stroboscopicly illuminated with a certain frequency. At low frequency, the observer perceives a single object. As frequency increases, the observer suddenly sees two objects, because the percept created by the first flash is still perceived when the second flash occurs. The lowest flash frequency at which 2 objects are perceived is used to determine the duration of persistence. (e.g. Dixon and Hammond 1972)

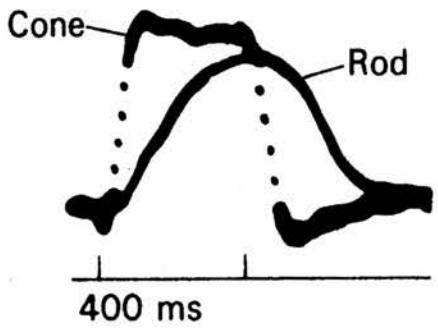
(6) Viewing through a moving slit. Imagine a stimulus (e.g. a square), and above it a occluder with a narrow vertical slit. If the slit is moved back and forth over the stimulus, the whole stimulus will suddenly becomes visible as a whole, through a

phenomenon sometimes called 'retinal painting', i.e. the individual photoreceptor receive their input at different times, but through persistence, the parts form a whole. In a way, the same is true for a computer screen: at 72 Hz, the top line is presented 14ms before the bottom line, yet through neuronal and phosphor persistence, they form a whole stimulus. The same is also perceived, when one drives by a picket fence: at low speed, one sees only part of the scene through the fence. At higher speeds, one perceived the entire scene, as if the fence had become a transparent surface. The lowest sweep rate resulting in perception of the whole object is used to estimate persistence.

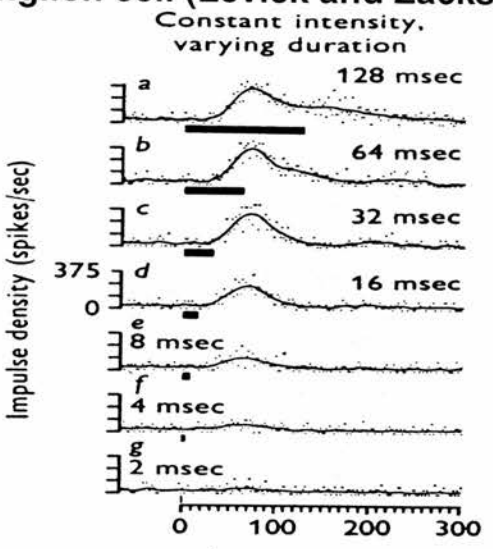
(7) Stereoscopic persistence. Random dot patterns are presented to the two eyes. The visual systems fuses them and a shape becomes recognisable, floating over the background. Using this paradigm, both (i) monocular and (ii) binocular persistence can be measured. One can present (i) one pattern to one eye, and the other to the second eye after a certain ISI. The longest ISI at which the shape can be correctly perceived is an estimate of monoptic persistence. Alternatively, (ii) a variant of the phenomenal continuity method seen in (3) can be used to assess the persistence of the percept that emerges from the binocular fusion: the 2 stimuli are presented simultaneously to the two eyes, but flickered on and off. The longest ISI at which the flickering pattern percept seems never to disappear is then used. Engel (1970) showed that this second form of persistence last for up to 300ms. Since the shape that is perceived can only be extracted at cortical level, after the two ocular channels have fused, this form of persistence demonstrates the existence of purely cortical neural persistence exists. Interestingly, the monocular persistence as measured by method (i) is shorter than that found in (ii) suggesting that cortical persistence can be longer than the pre-cortical persistence it receives input from.

Neural Persistence Overview

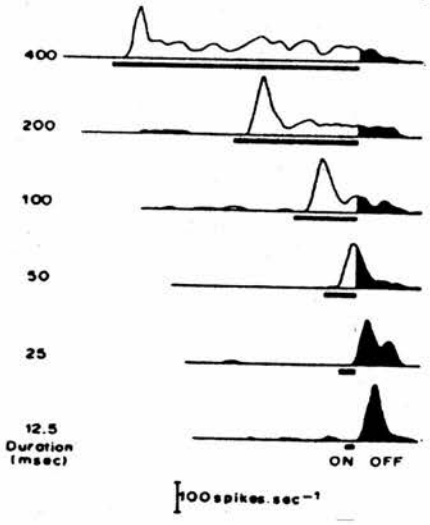
(a) Receptor (Whitten and Brown 1973a)



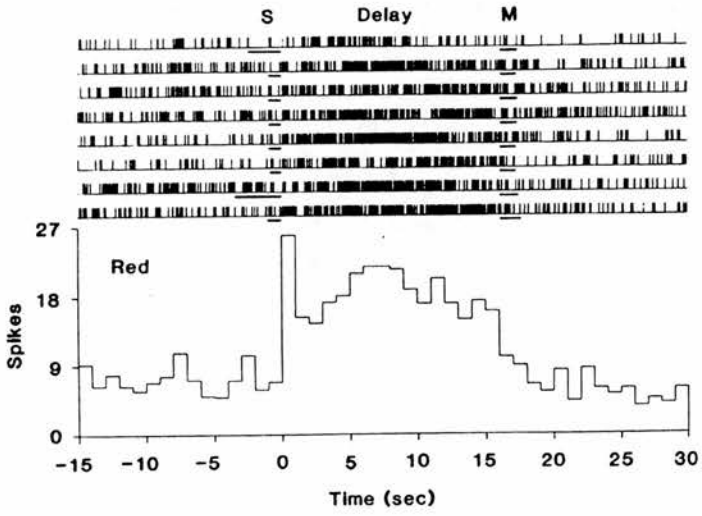
(b) Ganglion cell (Levick and Zacks, 1970)



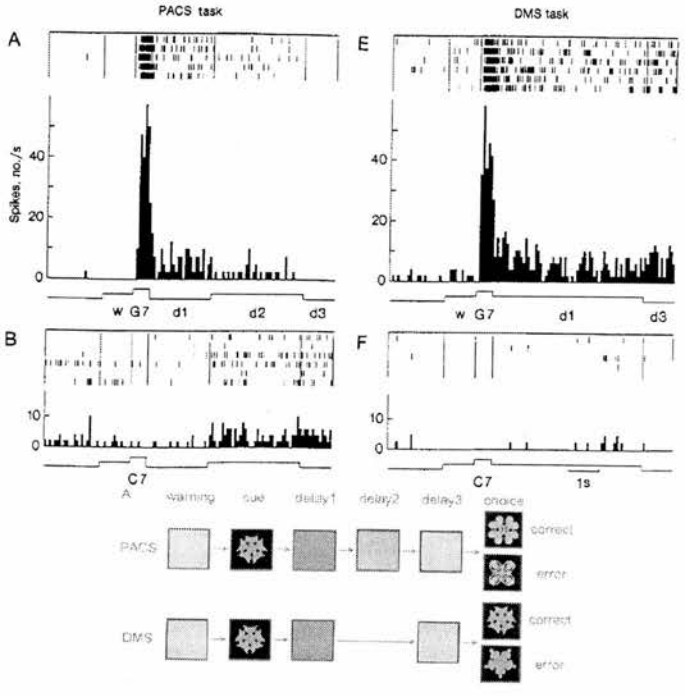
(c) Area 17 (Duysens et al. 1985)



(d) IT (Fuster, 1981)



(e) AIT (Naya et al. 1996)



(f) IT+STS (Rolls et al. 1999)

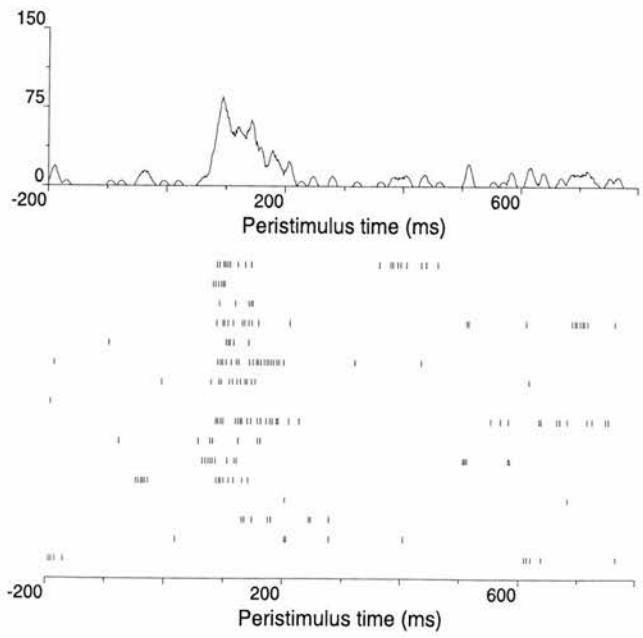


Figure 3.2: (Previous page) Overview of single cell evidences for neural persistence at different stages of the visual system (as indicated) adapted from the mentioned sources. (a) Extracellularly measured late receptor potential of a cone and a rod receptor in the macaque retina to a circular 400ms flash (as indicated by the tick marks) in the centre of the receptive field. (b) Spike density functions (lines) and psth for a ganglion on-centre cell in the anaesthetised cat as a function of flash duration (while stimulus intensity is kept constant). (c) Spike density function of a single neurone in Area 17 of the anaesthetised cat to an optimised stimulus of variable duration, as indicated by the black bar and numbers on the graph. In black, the area of neural persistence is indicated. To be comparable to visible persistence, rather than starting at stimulus off-set, the black area should start at stimulus off-set + response latency. (d) Rastergram and PSTH for a single neurone in IT responding to a red dot during a delayed matching to sample task. S indicates occurrence of the sample stimulus – the red dot - and M the moment, 18s later, at which 2 stimuli are presented for the monkey to choose S from. The duration of S, as indicated by the first bar in the rastergram, is variable, because the monkey terminated S presentation by pressing a button, when he was ready. (e) Comparison of the response of a single AIT neurone during a pair associate task (PACS) and a delayed matching to sample (DMS) task, for its optimal stimulus G7 and the ineffective stimulus C7, which was the pair associate of G7. Underneath, an example of a sequence of events and the correct choice for the two tasks (original in colour). During delay 1 (green screen), the monkey expects a DMS task. Occurrence of delay 2, i.e. a switch to a turquoise background, indicates to the monkey that this trial is a PACS trial. Delay 3 (grey screen) indicates the upcoming choice between two stimuli. (f) Spike density function (top) and rastergram (bottom) for a single neuron in STS to a 16ms, unmasked, presentation of the most effective face for that cell as a function of peristimulus time relative to stimulus onset. See text for details.

Neural persistence, i.e. the fact that the neural activity produced by a stimulus outlasts the stimulus presentation time, exists throughout the visual system. Figure 3.2 gives an overview of some important single cell studies indicating neural persistence. As early as in the **single photoreceptors** (Fig. 3.2a), the response to a flash of light outlasts stimulus duration, especially in the rod receptors. Interestingly, the response of single photoreceptors shows the inverse background intensity effect that is also present in its perceptual correlate, the visible persistence: Whitten and Brown (1973) could show that a bright 320ms flash of light produces a receptor potential in the cone receptor of the monkey that can last for several seconds if the monkey is fully dark adapted, while lasting only a few more milliseconds than the stimulus if the monkey has been pre-exposed to light for 120s. Since all further stages of the visual system depend on the output of the photoreceptors, it is no wonder, that Levick and Sacks

(1970), recording from on-centre **retinal ganglion cells** (Fig. 3.2b) in the anaesthetised cats, showed that however short a small disk of light is flashed to the centre of a receptive field, the response in the ganglion cell will never last for less than ~60ms, and will generally outlast stimulus duration even for stimuli longer than this critical minimum response duration. At the **first cortical level**, area 17 of the cat, Dysens et al. (1985) recorded neural persistence of about 50-100ms (Fig. 3.2c), with inverse duration and intensity effects apparent in some of the cells. At **higher cortical level**, neural persistence is maintained, with clear response durations of 100ms for faces presented for 16ms without a mask in the STS (Rolls et al., 1994, 1999; Rolls and Tovee, 1994; Fig. 3.2f). In Rolls et al.'s task, the monkey had just to fixate the screen, and had no reason to actively remember the stimulus.

If on the other hand a monkey is required to remember a stimulus to fulfil the requirements of a task, a different and much longer kind of persistence can be observed. Fuster and Jervey (1981) used a typical delayed matching to sample (DMS) task, in which the monkey sees a single stimulus (e.g. a red dot). Then after a delay period of blank screen (18s in their task), the monkey has to choose the matching stimulus out of a choice of simultaneously presented stimuli. To choose the right stimulus after the delay, the monkey has to keep the representation of the first stimulus active. The maintenance of stimulus representation can be observed neurophysiologically as 'delay activity'. As can be seen in the Fig. 3.2d, the presentation of the cell's preferred sample stimulus (S) results in a strong response, that persists to a lower extent throughout the 18s delay interval while the stimulus is no longer on the screen. The delayed activity is terminated after the presentation of the stimulus to be matched (M). This activity represents a correlate of short-term memory, or working memory, which is dependent on the mnemonic nature of the task and hence is different from the simple neural persistence that occurs in the absence of a mnemonic task (as used by Rolls and co-workers).

The fact that the long delay activity observed in DMS tasks is really dependent on the demands of the task becomes apparent in the experiment of Naya et al. (1996). They trained their monkeys not only on the DMS task, but also on a pair associate task (PACS). In this second task, the monkey had learned to associate a number of visual symbols with other visual symbols. The two tasks differed in their requirements: If symbol G7 was presented, in the DMS task, G7 had to be chosen later, while in the PACS task, C7, the pair associate of G7 has to be chosen. In their

very elegant study, they could therefore directly compare the activity during the delay interval as a function of task-requirements. Intuitively, one would expect, that a neuron firing to G7, should continue to fire during the delay interval in a DMS task, but not in the PACS task, since in the latter, C7, an ineffective stimulus for the neuron had to be chosen after the delay. On the other hand, if C7 was presented, one would expect delay activity in the PACS but not in the DMS task. As illustrated in Fig 3.2e, this was exactly what they found in some neurones. During their delay interval, the colour of the screen could either remain green throughout, in which case the monkey knew he had to perform a DMS task. In this case the neurone showed a prolonged delay activity throughout the delay phase if G7 was presented (Fig. 2e part E) but not if C7 was presented (part F). On some of the trials, the background colour changed from green to turquoise midway through the delay interval, indicating that after the change (d2) a PACS task and not a DMS task had to be performed. Accordingly, during d2, the neurone fired if C7 was presented previously (part B) and not G7 (part A), indicating the retrieval of G7 as the pair associated of C7 into working memory.

In both the studies of Fuster (1981) and Naya et al. (1996), a form of prolonged neural activity is displayed, that is dependent on a mnemonic requirement, and persists for many seconds, and is thus different from the automatically occurring neural persistence occurring in the same cortical levels in the absence of an explicit memory task. Very long responses occuring after the offset of a stimulus have also been observed after a stimulus disappears behind an occluding screen (Baker, Keysers and Perrett, 1998). In the latter case, there is no explicit memory task, but it is likely that throughout his life, the monkey has learned that it is a good idea to actively keep track of where is who, especially if "who" is a human experimenter that might go around giving food.

Altogether, it therefore appears, that neural persistence exists at all levels of the visual system. Neural activity seems to outlast stimulus duration by about 60-150ms in the absence of a mnemonic task. As we will see later, presentation of a mask after the stimulus can interrupt this neural persistence, just like it can interrupt the visible persistence of the stimulus. The fact that neural persistence occurs as early as the retina single photoreceptor indicates that it originates in the very first step of visual processing. All subsequent stages base their input on these photoreceptors. It is hence no surprise that visible persistence is carried over to those later stages. Nevertheless this does not show that later stages of processing do not contribute to an

increase of neural persistence, nor does it mean, that later stages are completely incapable of separating events occurring within ~60ms of each other. The contribution of later stages to persistence is indicated by the occurrence of stereoscopic persistence (see above), which is necessarily the product of a processing stage above the point of binocular fusion, i.e. a cortical processing stage. The fact that later stages may overcome the seemingly limiting factor of early neural persistence is indicated by studies presented in the following section.

3.1.2.1. Neural persistence and the resolution of rapidly presented stimuli.

If single photoreceptors show neural persistence in the order of 60ms, one might think, that like a camera with an opening time of 60ms, the retina would be unable to resolve events occurring within those 60ms. That is to say, if two brief flashes F1 and F2 are presented within 60ms of each other, they would fuse into a single response, and be treated as a single stimulus. At first glance, this assumption seems reasonable. At the level of single ganglion cells, it has actually been empirically confirmed.

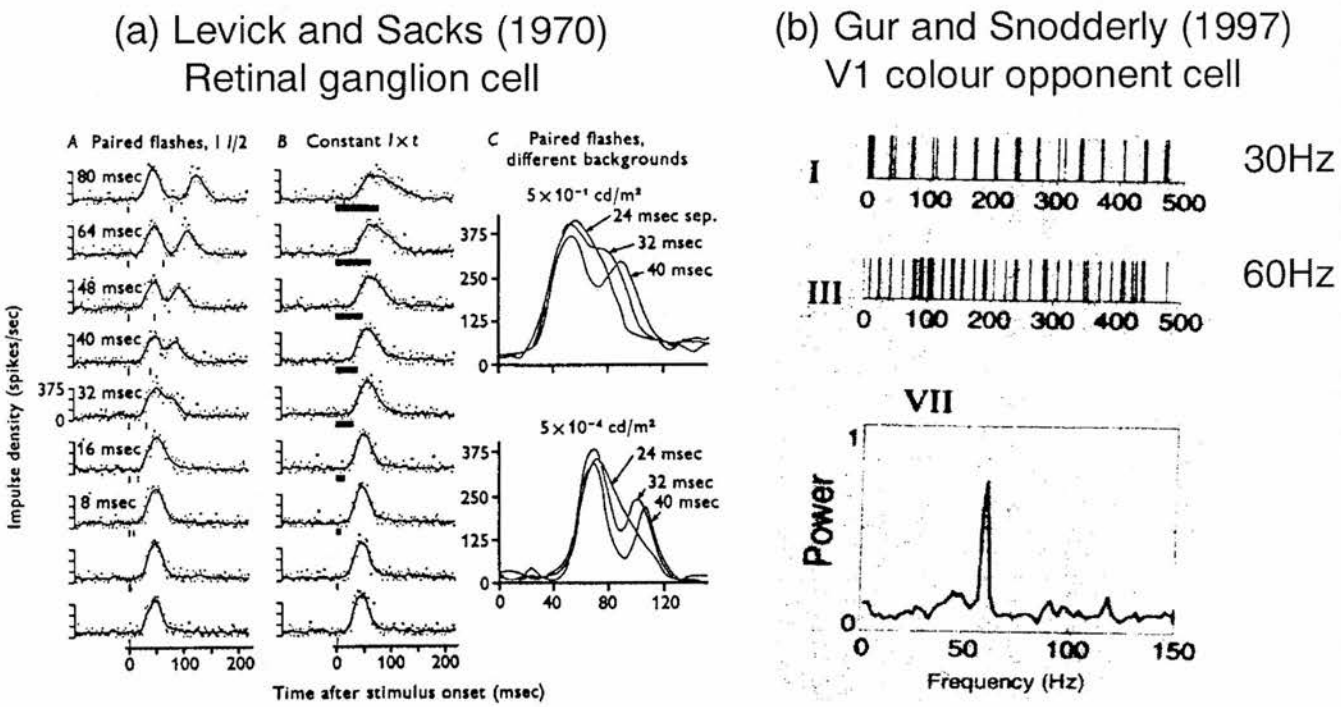


Figure 3.3: Temporal resolution in the retina's ganglion cells and V1. (a) the spike density function (line, smoothed using a 6ms Gaussian) and PSTH (dots) of the response of an on-centre ganglion cell to (A) a pair of flashes (2ms, 21cd/m^2 on a 0.5cd/m^2 background), or (B) to a single flash of identical total energy and lasting from the onset on the first flash to the offset of the second flash in (A). As can be

seen, the response to the pair of flash starts fusing at SOA=48ms, and only a single peak remains apparent in the spike density function at SOA=16ms, at which point the response becomes identical to the single flash of 18ms duration. (C) Integration is more pronounced at *higher* background illuminations, a surprising effect given the general inverse background illumination effect of persistence. At both illuminations, two flash at SOA=24ms produce a single peaked response. (b) I. Response of a V1 colour opponent (B+/R-) cell to a blue bar flashed on and off at 30Hz on an isoluminent red background. III. Another V1 cell (R+/B-) responding to a "steady" 8cd/m^2 red bar on an isoluminent grey background on a 60Hz monitor. VII. Power spectrum of the cell in III. Note the clear peak at 60Hz in VII and the 6 bursts every 100ms in III, indicating that the cell was able to resolve flashes occurring with SOA=16.7ms. It is striking to see how different the response to SOA=16ms flashes are in the ganglion cell and in V1. The difference is unlikely to be explained by stimulus intensity, as the stimulus in (a) is 3 times brighter than in (b), and a brighter stimulus results in a shorter persistence. Possibly, the difference in background intensity (0.5cd/m2 in (a) and 8cd/m2 in (b)) may account for the difference, given the inverse background intensity effect observable in Whitten and Brown (1973).

Levick and Sacks (1970) recorded the responses of single on-centre retinal ganglion cells in the anaesthetised cats (Fig. 3.3a). They showed, as discussed above and illustrated in Fig. 3.2b, that how ever short a small disk of light is flashed to the centre of the receptive field, the response in the retina will never last for less than ~60ms. But they also show, that if two brief flashes of 2ms duration are shown one after another, the responses they produce start to overlap at SOA<60ms and fuses into a single response at SOA=24ms. It should be noted, that the 6ms Gaussian they used for smoothing will contribute to erase bimodalities in the actual response. Inspecting the PSTH itself (dots) indicates that they might still be a small bimodal component even at those shorter SOA. The single fused response is nevertheless almost identical to the response produced by a stimulus lasting from onset of the first to offset of the second flash and having the same energy. Hence, as far as a single ganglion cell is concerned, the two stimuli are integrated into a single representation that is very similar to the presentation of a single stimulus of equal energy.

From Levick and Sacks (1970) work, one would assume that flashing a point on and off with 24ms (~40Hz) ISI should result in fully fused responses. Nevertheless colour opponent ganglion cells in the macaque show time locked response modulation over time with colour flicker at 40Hz, i.e. when a red dot is shown on one frame and a green dot on the next frame of a 80Hz monitor (Lee et al., 1989).

What happens when a optimal stimulus is presented continuously on a 60Hz monitor? Although it appears, that the stimulus is continuously presented, this is physically not true. A single frame of a 60Hz monitor is drawn in 16.7ms, with the cathode-ray drawing the page from upper left to lower right corner. A stimulus occupying $1/10^{th}$ of the height of the screen is thus drawn in 1.6ms, and given a 1ms phosphor decay time typical for a standard monitor, the stimulus is gone 2.6ms after it started, followed by a pause of 14ms during which the monitor is black. Gur and Snodderly (1997) (Fig. 3.3b) could show that many (but not all) neurones in LGN and V1 reflect the fact that their optimal stimulus on a 60Hz monitor actually goes on and off 60 times a second: the neurones show 6 bursts every 100ms, and a clear 60Hz peak in the Fourier Frequency Spectrum of their response. Hence it appears, that single cells in V1 and LGN can resolve the gap between stimuli presented with 16.7ms SOA: a surprising finding given the apparent fusion of the retinal responses of stimuli presented at such SOA in Levick and Sacks (1970). This apparent discrepancy might be due to a difference in species (cat/monkey), in stimulus parameters (background illumination etc.), in methods of analysis (smooting function, averaging of trials with latency jitter), or it may be that the surprising discrimination power of LGN and V1 cells may result from a dynamic interaction of on and off-centre ganglion cells. Interestingly, the neuronal capacity to resolve 60Hz flicker in Gur and Snodderly (1997) is in contrast with the fact that at 60Hz, human subject do not perceive flicker in central vision (where the neurones of Gur and Snodderly have their receptive fields), suggesting that V1 activity is not always accessible to consciousness.

Altogether while persistence does occur for about 60ms in the retina and does result in integration of temporally neighbouring events, the response of some V1 cells clearly indicates that flashed with SOA=16.6ms are two and not one flash. This effect shows the necessity to investigate the neural representation of short events in later visual areas, since it is misleading to believe that single cells in the later visual system can only be temporally sloppier than individual neurones in earlier levels.

3.1.3. Information persistence

Information persistence is a more abstract form of persistence than visible persistence, and is best understood by its operational definition through a paradigm called partial report which has been introduced by Sperling (1960), and is still used in basically the

same fashion (e.g. Loftus and Irwin, 1998). A matrix of letters, e.g. 3 rows of 4 letters, is presented briefly (<500ms), and subjects are asked to report all the letters they have seen. In that 'full report' condition, subjects report on average 4 to 5 out of the possible 12 items. Sperling's innovation was to present the same matrix, and after an ISI he gave a tone indicating which row had to be reported (high pitch = top row, middle pitch= middle row etc). Naively, one would expect, that if the cue is given after the stimulus has been turned of, the subjects could only apply the cue to the 4 or 5 stimuli they had transferred to short-term memory, as revealed by the full report condition. On average they should therefore only be able to report 4.5/3=1.5 items. Surprisingly, Sperling found, that for a matrix presentation of 50ms, and an ISI of less than 300ms, the subject reported more correct items, than statistically predicted by their full report performance – a phenomenon called 'partial report advantage'. They typically report 3 out of the possible 4 letters of the cued line, while they would only report 4 letters out of the possible 12 in the full report condition. Sperling interpreted this to mean, that there is a limit transfer capacity from the image to the working memory, which was of only about 4 items. If no cue was given, these 4 items were taken at random from the 12 in the matrix. If a cue was given, they could be taken selectively from the right row, resulting in a much better performance.

In addition, there is an iconic memory (so called by Neisser, 1967), which keeps the image alive in a pre-categorical form akin to the original stimulus for a few hundred milliseconds beyond the stimulus physical duration. The transfer from image to the more durable working memory could hence be continued, or selectively directed, after the stimulus was turned off, but while the iconic memory was still alive. The iconic memory decays very quickly, and is gone by about 300ms, as measured by the longest ISI between matrix and cue, at which a partial report advantage is observed.

Till about 1980, this form of iconic memory was taken to be equal to visible persistence. Then, Coltheart (1980a,b) urged that equation to be questioned, and differentiated between visible persistence and information persistence (i.e. iconic memory). Amongst others, a recent paper by Irwin and Loftus (1998) explicitly compares information and visible persistence, and confirmed the difference. (Fig. 3.4). In particular, visible persistence, as measure by their subjective report task (Fig. 3.4b) has a shorter duration (only about 60-100ms) than information persistence (Fig.

3.4a). Information persistence does not seem to show the inverse duration effect so typical of visible persistence. It turns out that this is true only if image intensity is kept constant, and only duration is varied. If on the other hand, intensity is corrected so that stimulus energy (i.e. duration x intensity, Bloch's Law) remains constant, then information persistence too shows the inverse duration effect (e.g. DiLollo and Dixon, 1992, and experiment 4-7 of Irwin and Loftus, 1998).

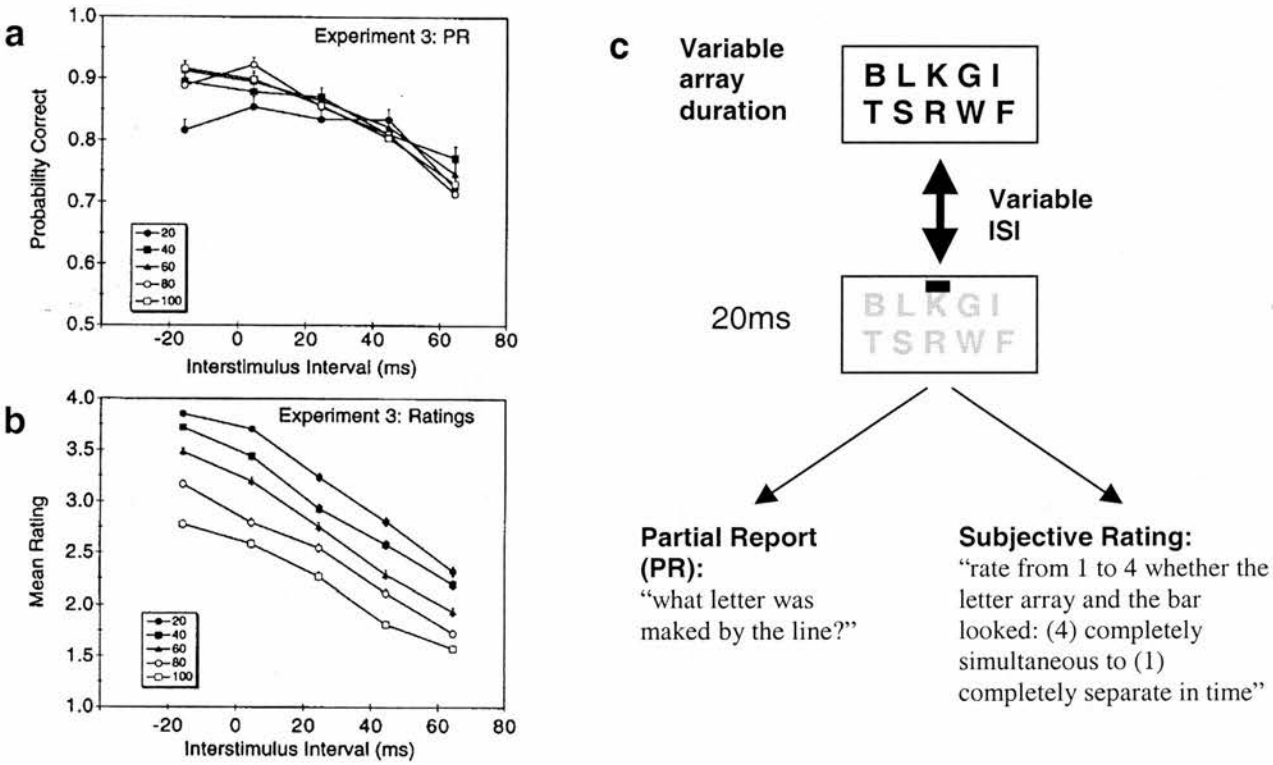


Figure 3.4 (c) Experimental design in the study of Loftus and Irwin (1998). A 2x5 matrix of consonants was flashed, followed after an interstimulus interval (ISI) by a 20ms bar (the light grey letters were not presented - they symbolise the decaying icon of the letter matrix). Depending on the block the subject was in, he had either to report which letter was indicated by the bar (partial report), or how simultaneously the bar and the matrix was perceived, ranging from 4, fully simultaneous to 1 fully separate. The results of these two tasks as a function of ISI (x-axis) and matrix duration (different lines) are shown in **(a)** and **(b)** respectively. Note the inverse duration effect in **(b)** but not in **(a)** [Adapted from Loftus and Irwin, 1998]

Loftus, Duncan and Gehrig (1992) investigated the effect of a visual noise mask on iconic memory. Their task was somewhat different: they presented 4 digits for a variable amount of time, and asked subjects to report the digits immediately. In addition, they presented a visual noise mask at variable ISI after the digits, and

measured the effect of ISI on number of reported digits. They found that the number of digits correctly reported increases with ISI. They interpreted this as indicating, that (i) there is an iconic memory of the digits, that permits extraction of information about the digits after the images has been terminated and that (ii) a mask can erase or interupt this iconic memory. Based on these ideas, they defined the time course of the decay of the iconic memory from which digits are extracted as long as no mask is presented, and found this function to be independent from stimulus duration (unlike visible persistence), and to follow a gamma function.

3.1.3.1. What is stored in information persistence?

Information persistence is more abstract than visible persistence, because it is not associated with our direct perception of the stimulus, as shown by Irwin and Loftus (1998). Nevertheless the information stored in information persistence must be of a nature resembling the stimulus itself, since cues such as spatial position can be used in partial report cuing. These cues would be ineffective, if only the semantic content of the image was stored. Nevertheless, findings that will not be discussed here in detail (e.g. Smythe and Finkel, 1974; Eriksen & Rohrbaugh, 1970; Dick, 1974), indicate that spatial location and stimulus identity are stored separately in iconic memory, with the spatial information decaying very rapidly (about 90ms), while the identity can persist somewhat longer (about 240ms). This results in the fact, that at intervals between 90ms and 200ms, if asked to report a particular letter, errors will often be a neighbouring letter, since the identity of that letter will still be in iconic memory, while its precise location has already faded away. Other studies also indicate, that post-categorical[4] information can affect the way iconic memory is used. For instance, Duncan 1982 showed, that asking subject to report the digits but not the letters from an alphanumeric matrix results in a partial report advantage. This is puzzling, because it conflicts with earlier findings of Sperling (1960) and von Wright (1968), and because it means, that the selective mechanism transferring information

[4] The distinction "pre-categorical" and "post-categorical", although popular in the psychological literature is not supported by our knowledge of brain physiology: a V1 neurone for instance, detecting a certain bar of a certain orientation, categorises e part of a face as "a line of 90deg orientation", a STSa neurone responding to a face of a certain orientation categorises the same stimulus along different categories "a profile face". Both are post-categorical, but at different levels of categories. A hard distinction "pre-categorical", "post-categorical" is thus artificial, and not supported by the physiology of the brain, which appears to performs a continuous categorisation along increasingly "high level" categories.

from the original stimulus to the working memory, must have access to post-categorical information (e.g. is it a letter or a number?).

Based on these findings, Coltheart (1983) revised his model of memory to include three stages of memory, which - using slightly different words - could be called: pre-categorical iconic memory, a post-categorical iconic memory and working memory. There is in addition a transfer mechanism, of limited capacity, that can transfer about 4 items from either one of the rapidly fading iconic memories into the longer lived working memory.

3.1.4. Conclusions

Turning a stimulus off on a screen does not turn the stimulus off in the mind or in the brain. Three distinct classes of phenomena have to be distinguished. Visible persistence refers to the fact, that our perception of a stimulus outlasts the stimulus itself for a few hundred milliseconds. During that time, the stimulus progressively fades away from our mental eye. Information persistence refers to the fact, that if asked to report the identity of a number of stimuli, a subject can continue extracting information about the stimulus after it has been turned of. Neural persistence, on the other hand, refers to the fact that the neural activity induced by a stimulus outlast the stimulus.

3.1.4.1. Relationship between neural persistence, information persistence and visible persistence.

From the perspective of a neurophysiologist, neural persistence is the least mysterious form of persistence. It can be measured throughout the visual system. What is interesting about it, is that the information represented in the different stages of processing is very different. At the retina, the LGN or V1, the information is very simple, and reflects the existence of 'pixels' or simple contours in the stimulus,. This information is precategorical: single cells at those levels do not explicitly represent a certain category such as "it is a face", or "it is a letter". Hence, it is tempting to associate this earlier activity with more graphic, pre-categorical aspects of persistence. Neural persistence in later areas, such as STS represent complex shapes, and generalise over members of a category. Such activity is more likely to be associated with a more post-categorical form of information persistence, such as the one revealed by Duncan (1982). Interestingly, the fact that information about location and identity

fade at different rates (e.g. Smythe and Finkel, 1974) makes neuropsychological sense: if neurones in the ventral stream, such as those in STS represent the identity of an object, their receptive fields are large, and will only give coarse information about the position of the stimulus (e.g. Baker, Keysers et al., In press). Other neurones, with smaller receptive fields, would give precise information about location, but not identity. Hence, if those different stages show different durations of persistence, then information about identity and location would fade at different rates.

The link of neural persistence and visible persistence is more difficult. Our understanding of how neural activity correlates with the subjective perception that visible persistence measures remains embryonic at best. And the question of which neural persistence participates in visible persistence is almost identical to asking which neural activity is conscious, and is thus equally unanswered.

3.2. Masking

Visual masking is the interference caused by one stimulus on the perception of another spatio-temporally overlapping or adjacent visual stimulus. The term visual masking ("masquage visuel") is attributed to Piéron (1925), while the study of the phenomenon is much older (see Breitmeyer, 1984, p1-26 for a historical review).
The basic masking paradigm uses two stimuli following each other shortly in time. The amount of perception of one of the stimuli, the "test" or "target" stimulus is then measured, while the second, "masking" or "conditioning" stimulus serves to disturb that perception. The aim of the investigation is to measure the effect of variations of the temporal and spatial arrangement and nature of the two stimuli on the measure of perception of the target stimulus. The phenomenon of visual masking has received much attention in psychology, and an extensive review would go beyond the scope of this thesis. There are many excellent reviews of the topic published in journals (Felsten & Wasserman, 1980; Breitmeyer & Ganz, 1976; Turvey, 1973; Kahneman, 1968) and books (Breitmeyer, 1984; Humphreys & Bruce, 1989). In this section, a brief review of the central findings in the investigation of masking will be presented, together with the first attempts to find physiological correlates of this effect. Masking is particularly relevant to this thesis, because RSVP, the paradigm used in our experiments is an extension of pattern masking to cases when a target pattern is not masked by one but by many stimuli, all of which could themselves be targets.

3.2.1. Terminology in the study of masking

Depending on the nature and spatio-temporal relationship of the stimuli, the following types of masking are distinguished (Fig. 3.5).

If the "test" or "target" stimulus (TS) is fully contained in the area of presentation of the "masking" or "conditioning" stimulus (MS), three types of masking are distinguished based on the nature of the MS. In **masking by light** the TS is a small flash of light, or a pattern such as a letter, and the MS is a homogeneous field of light typically much larger than the TS to avoid any interactions between the contours of the stimuli. In **masking by pattern**, the TS is a pattern such as a letter, and the MS is a non-homogeneous field composed of patterns such as letters, lines

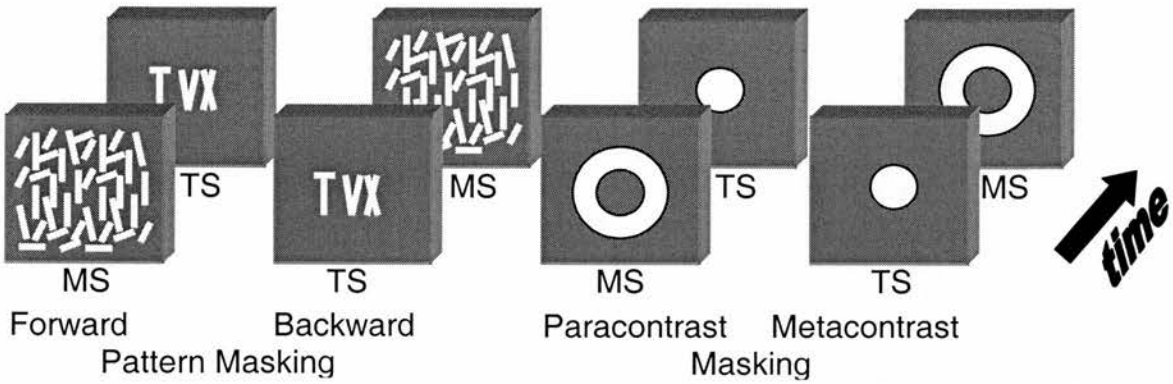


Figure 3.4: Arrangements of typical targets and masks in masking paradigms. TS indicates the target stimulus, and MS the masking stimulus. In the two tasks on the left, the observer's task is to name the 3 letters. In the two right tasks, the subject judges the luminance of the TS disk.

etc. In **masking by noise**, a special type of masking by pattern, the MS is a visual noise pattern, akin to the "snow" on an untuned television set. In those three cases, if the TS starts before the MS one speaks of **backward masking**, and if the MS starts before the TS one speaks of **forward masking**.

If the TS and the MS do not overlap in space, but are adjacent to one another, one speaks of **metacontrast masking** if the TS starts before the MS and **paracontrast masking** if the MS precedes the TS. Stimuli in metacontrast masking are typically a bar as TS and 2 bars flanking the TS bar as MS, or alternatively, a disk as TS, and an annulus concentrically surrounding the disk as MS.

The dependent variable in masking paradigms is usually either a probability of correct identification of the TS or a detection threshold (i.e. the minimum TS duration or intensity necessary for correct identification). The independent variables include the temporal offset between TS and MS, or their respective durations, as well as other stimulus parameter such as intensity, complexity, similarity, contrast, spatial frequency etc.

Two measures of the temporal relation of the two stimuli relative to each other are used: the interstimulus interval (ISI) or the stimulus onset asynchrony (SOA). As illustrated in Fig. 3.6a, ISI refers to the time from the offset of the first stimulus to the onset of the next stimulus, while SOA refers to the time between onset of the first and onset of the second stimulus.

Masking has either a Type-I (also called Type-A or monotonic) masking curve, in which case, the strongest masking and hence the worst perceptual

performance occurs at SOA=0; or it has a Type-II (also called Type-B, U-shaped or W shaped) masking curve, with little masking at SOA=0ms, strong masking for intermediate SOA (e.g. 60ms), and no masking again at large SOA (See Fig 3.6b)

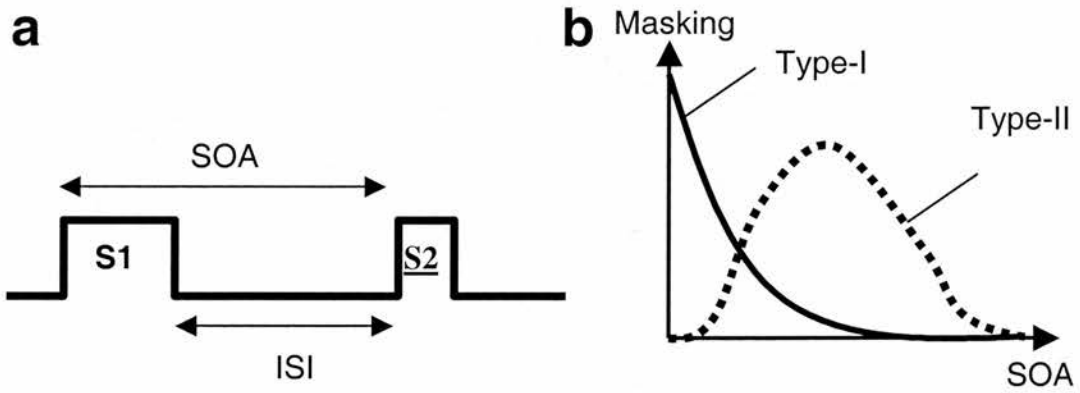


Figure 3.6: (a) Nomenclature in use for describing the temporal relationship between two stimuli S1 and S2. The thick line describes the onset and offset of two stimuli. The stimulus onset asynchrony (SOA) then describes the time elapsing between the onset of S1 and the onset of S2, while the interstimulus interval (ISI) describes the time elapsing between the end of S1, and the beginning of S2. **(b)** Type-I vs Type-II masking.

3.2.2. Theoretical framework

The phenomenon of persistence was reviewed earlier. Masking is intimately linked with persistence. Since in masking, TS and MS are usually not presented simultaneously, what interact are not the two stimuli themselves, but one stimulus with the persistence of the other, earlier stimulus. This interaction is of generally two types. The later stimulus can **integrate** with the persistence of the first, or it can **interrupt** the persistence of the first. Before getting into the literature, these two possibilities will be clarified. They do not necessarily exclude each other, and many a masking phenomenon is the combination of both.

Integration: refers to the fact, that two stimuli presented at ISI of less than ~60ms will be integrated to some extent by the visual system in a fashion that resembles the double exposition of a photographic plate. This is due to the temporal persistence of the first stimulus. Physiologically, integration is present as early as the retina. Levick and Sacks (1970), recording from on-centre retinal ganglion cells in the anaesthetised cats showed that however short a small disk of light presented to the centre of the

receptive field is flashed, the response in the retina will never last for less than ~60ms. If two brief flashes of 2ms duration are shown one after another, the responses they produce fuses or integrate in a single response for ISI of up to 30-50ms (depending on the criterion used). This single response is identical to the response produced by a stimulus lasting from onset of the first to offset of the second flash and having the same total energy. Hence, as far as a ganglion cell is concerned, the two stimuli are integrated and are indistinguishable from the presentation of a single stimulus produced by the integration of both stimuli. Levick and Sacks (1970) showed two other important findings. First, as the duration of a flash is reduced from 128ms to 2ms, the duration of the response is progressively reduced for stimulus durations of 128 to 32ms, without affecting the peak firing rate. For those stimulus durations, the response outlasts stimulus duration by ~60ms. Reducing stimulus duration below 32ms stimulus duration, the response always lasts for ~60ms, but the amplitude of the peak firing rate decreases. This seems to indicate two different temporal dynamics in the retina. For durations beyond 60ms, the retina can modify the *duration* of its responses. Below 60ms, it can not, and only the firing *rate or probability* can be altered, resulting in the integration temporally adjacent stimuli. Nevertheless, integration in the retina clearly occurs at SOA of 60ms or less, resulting, as with the double exposition of a photographic plate, in a masking due to a reduced visibility of the mutually camouflaging components.

Interruption: the presentation of a new stimulus interrupts the processing of the old stimulus. If the perception of a stimulus necessitates a certain amount of processing time, this interruption may prevent or reduce perception. For ISI over 60ms, integration plays little role, and interruption seems to be a rather pure mechanism. Under 60ms, the origin of masking is harder to isolate, and may result from both interruption and integration.

3.2.3. Psychophysical investigations of masking

3.2.3.1. Masking by Light

Masking by light occurs when a bright flash is used as a masking stimulus, and is generally measured as in increase in the minimal intensity necessary for the detection of a test flash.

The most cited study on masking by light is probably the classical study by Crawford (1947). The method employed by Crawford was to measure the detection threshold of a small, 10ms test flash centred and superimposed on a much larger conditioning field of variable luminosity that was turned on for 500ms (Fig 3.7). What makes this study particularly valuable is that by using a long conditioning or masking stimulus, the study could show, that the transient on- and off-set plays a particularly disruptive role, compared to the sustained part of the masking stimulus.

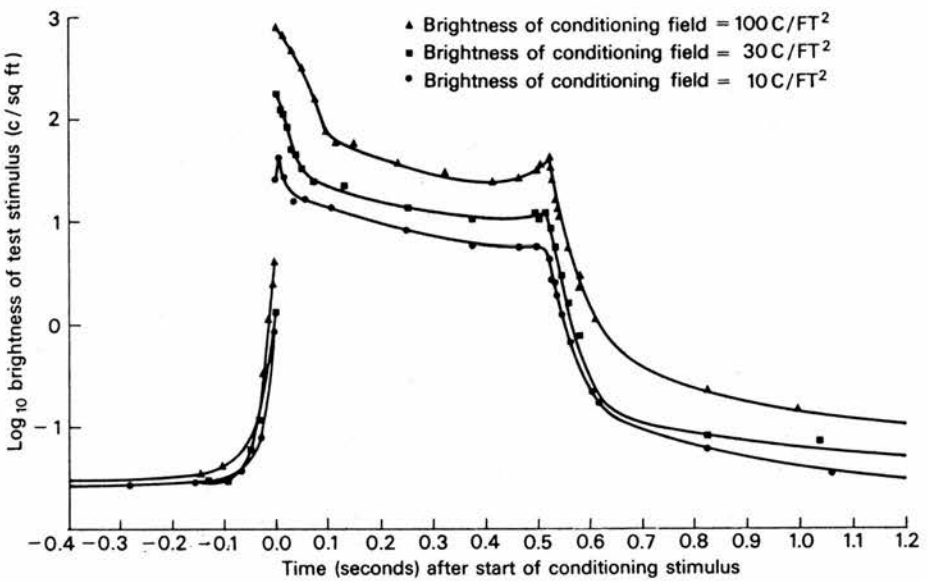


Figure 3.7: Log brightness of a test stimulus flash necessary for its detection as a function of the timing of the 10ms test flash relative to the start of the 500ms conditioning flash and the brightness of the conditioning flash. Note the overshoot of the detection threshold around 0ms and 500ms, i.e. during the transient of the conditioning flash, especially at high stimulus brightness. Note also, that the detection threshold starts increasing almost 100ms before conditioning stimulus start, and stays elevated many milliseconds after the conditioning stimulus end [Adapted from Crawford 1947].

The main findings of the study are: (i) The detection threshold for the target flash begins rising about 100ms *before* the start of the masking stimulus, hence turning on a mask 100ms before a target disrupts detection. (ii) The onset and offset of the masking stimulus produce overshoots (i.e. transient severe elevations) in the detection

threshold, especially at higher stimulus intensities, and (iii) after the end of the masking stimulus, detection threshold does decrease rather slowly, not returning to baseline for many hundreds of milliseconds. Point (i) and (iii) reflect backward and forward masking by light, showing that forward masking can be much longer than backward masking. Point (ii) is important because it draws attention to the special importance of the transient phases of a stimulus: both on- and off-transient produce a particularly strong masking, which led to an entire theory of masking as being the disruption of processing of one stimulus by the transient response to the next stimulus (see in particular Breitmeyer and Ganz, 1976; and Breitmeyer 1984 p162-232, but also Felsten and Wasserman, 1980; and Kahneman, 1968).

Interestingly, the transient overshoot on the Crawford function (see Fig. 3.7) seem to depend on the magnocellular system. Green (1981) also presented a 500ms MS, but his TS was a grating either of 1c/deg or 7.8c/deg. With the 1c/deg TS, the masking function showed the same transient overshoots at on- and offset of the mask. With a 7.8c/deg target this was not the case: the masking function showed a steady plateau during the entire MS duration without transients overshoots. This is taken as evidence, that the magnocellular pathway is responsible for the overshoot, and in particular, that the magnocelluler system inhibits itself (and hence the detection of the 1c/deg grating), more than it inhibits the parvo system (hence the lack of overshoot in the detection of the 7.8c/deg target grating).

Masking by a flash of light is mainly a monocular effect. If the TS and MS are presented to different eyes, the detection threshold for the TS remains as if no MS had been presented (e.g. Schiller, 1965), although minimal amounts of trans-ocular masking by light do occur. This indicates a predominantly pre-cortical locus for masking by light.

Generally forward masking occurs for SOA of –60ms to 0ms, and backward masking occurs for SOA of 0-200ms, although with more prolonged masking stimuli like the one used by Crawford, forward masking can be much longer, but this is probably due to receptor adaptation. Rise and fall time of the MS affect masking in predictable ways (Matsumura 1976a,b), reinforcing the idea, that brisk transients are particularly good masking stimuli.

While there would be much more to say about masking by light, masking by light is less relevant than masking by pattern to the later experiments in this thesis, and the interested reader should consult one of the many reviews on masking (Felsten

& Wasserman, 1980; Breitmeyer & Ganz, 1976; Turvey, 1973; Kahneman, 1968; Breitmeyer, 1984; Humphreys & Bruce, 1989) for further details.

3.2.3.2. Masking by pattern (and by visual noise)

Masking by pattern differs from masking by light, particularly because unlike masking by light, masking by pattern is strong also if the MS is presented to the eye contralateral to the TS, while nevertheless slightly weaker than ipsilateral masking (Schiller 1965).

Turvey (1973) performed a thourough investigation of masking by pattern. He presented a letter trigram (e.g. "VXM") and asked the subject to report the three letters. He then presented a mask in the temporal vicinity of the trigram, and measured how this affected performance (see Fig. 3.8).

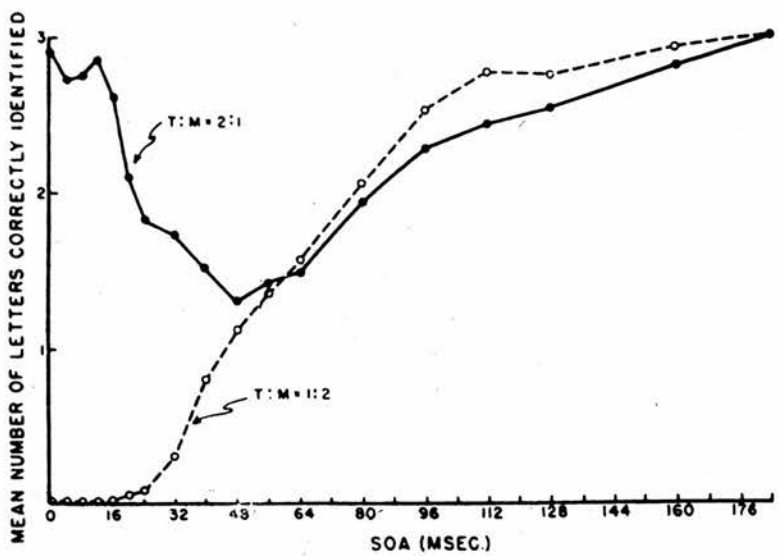


Figure 3.8: A letter trigram was presented for 10ms, followed by a 10ms pattern mask composed of features similar to those used in the letter trigrams. The number of correctly identified letters was measured as a function of SOA at two different energy ratios between Target (T) and Mask (M). [Adapted from Turvey, 1973, experiment XVIII].

The main finding was that presentation of a pattern mask up to about 160ms after the trigram can reduce the performance. Interestingly, during the first 60ms, the energy ratio of stimulus and mask played a critical role. If a target was stronger than a mask, masking followed a Type-II (also known as type-B or U-shaped) function: simultaneous target and mask presentation did not prevent identification of the target,

while increasing the interval up to about 48ms increasingly disrupted recognition. Thereafter, for SOA>48ms, performance increased again. The fact that simultaneous presentation of the mask did not camouflage the target (SOA=0) is the key to understanding this phenomenon (Fig. 3.10). It indicates that if target energy is larger than mask energy, even full integration of the two stimuli during simultaneous presentation did not interfere with performance. Hence, what ever integration occurs at that energy ratio will not be reflected in the identification performance. The drop of performance at SOA around 30-120ms thus cannot be accounted for by integration, and has to be understood as occurrences of stimulus competition or interruption. Stimulus competition means, that the representation of the target and the mask mutually disrupt each other due to inhibitory competition. Interruption is very similar, but indicates that the new stimulus wins the competition and interrupts the representation of the old stimulus.

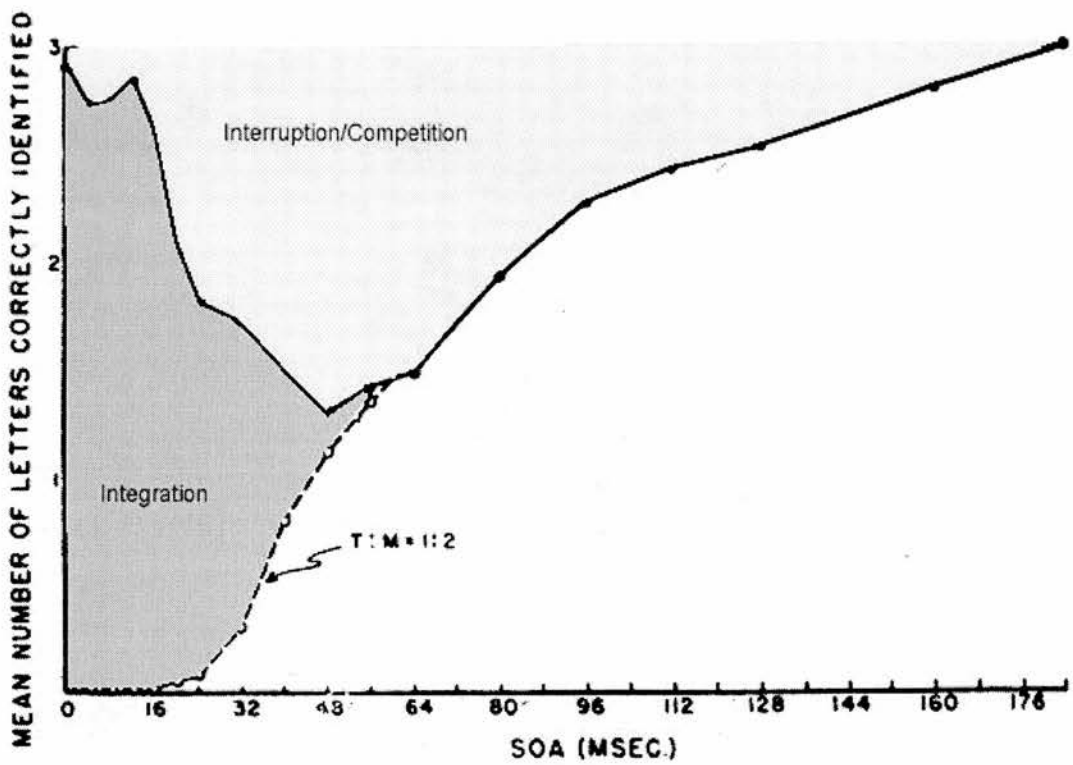
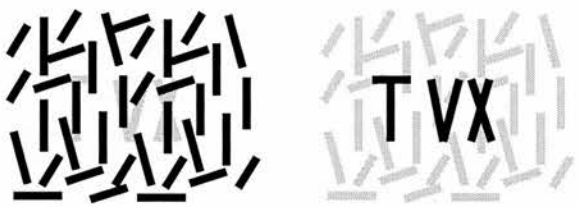


Figure 3.9: Hypothesised contribution of integration (dark grey) and interruption/competition (light grey) in the experiment of XVIII of Turvey (1973) if an additive combination of integration and interruption is assumed.

When the energy ratio is in favour of the mask, one can disclose the amount of integration occurring in the system: a Type-I or monotonic function will result, with

maximal masking at SOA=0 (Fig 3.8). In the light of that reasoning, one can obtain some estimate of integration and interruption/competition from the Turvey (1973) experiment. Since in cases where the energy ratio is in favour of the target, integration cannot disrupt performance and deviation from perfect performance can only reflect the effect of interruption or competition between the stimuli (light grey area in Fig. 3.9). Measured in that way, Interruption/Competition peaks at SOA=48ms, and disappears at SOA=176ms. If as a very first approximation one assumes that integration and interruption are additive effects, the amount of integration occurring in the case of energy ratios in favour of the mask may correspond to the impairment of performance above and beyond interruption (dark grey area in Fig. 3.9). Integration would then peak at SOA=0ms and vanish at SOA=64ms. Of course this estimate is very crude, since evidently integration will reduce the contrast of a stimulus, and hence reduce the amount of interruption the neural representations of the stimuli can produce, producing non-linear relationships between interruption and integration.



TS:MS=1:2 TS:MS=2:1

Figure 3.10: An artists impression of the integration of the target and the mask in Turvey (1973) given the target stimulus (TS) and masking stimulus (MS) energy. The task is to name the three letters of the TS. MS and TS were presented for 10ms at luminances of 2.5 or 5 footlamberts

Generally, monoptic forward masking by noise produces stronger masking effects, that extend to longer SOA than monoptic backward masking (Kinsbourne and Warrington, 1962a,b; Schiller, 1965; Schiller and Smith, 1965). Yet remarkably few studies have investigated the combined effect of forward and backward masking. One such study is that by Uttal (1969a,b). His stimuli were letters composed of dots, and presented for 0.4ms. Either before, after or both before and after the stimulus, he presented dynamic visual noise (DVN), i.e. during one seond, he presented single dots at random positions on the screen, at a rate of one dot every 1, 2 or 3 ms. Using this paradigm he could measure the effect on letter identification of DVN before, after or both before and after the stimulus.

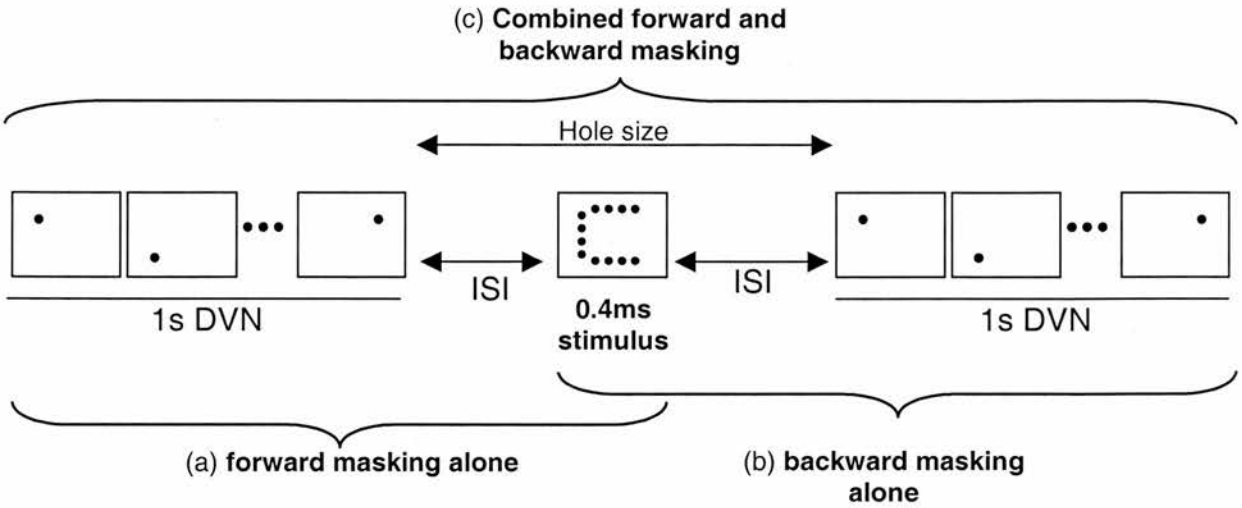


Figure 3.11: Experimental design in Uttal (1969a,b): (a) in the forward masking condition alone, 1s of dynamic visual noise (DVN) composed of single dots presented at random locations at a rate of one ever 1, 2 or 3ms is presented on the screen, followed by an ISI of blank screen of variable duration, and finally by a dot letter presented for 0.4ms. The task was to name the letter. In the backward condition (b) the letter was presented, followed by the DVN after a variable ISI. In the combined task (c) the letter was in a "hole", i.e., it was flanked by DVN before and after the stimulus.

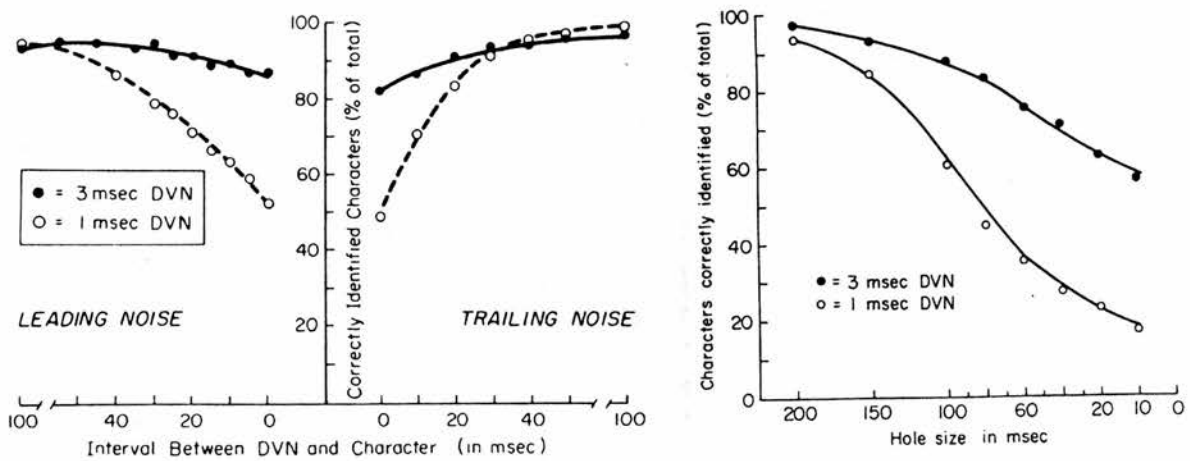


Figure 3.12: Effect of DVN forward and backward masking separately (left), and their combined effect (right) as a function of Interval between DVM and stimulus (left) or hole size between the two DVN (right), the stimulus being in the middle of the hole. The effect is shown separately for DVN presented one dot every 3 or one dot every 1ms. Note that given the persistence of each dot in the DVN, the DVN is perceived as being a mask containing about 30 dots or 100 dots for the 3ms and 1ms interdot interval respectively [Adapted from Uttal 1969b]

Using forward and backward masking separately, he confirmed that forward masking extended for slightly longer intervals (see Fig 3.12 left). But most interestingly, he found that using masks both before and after the stimulus (combined masking) resulted in stronger masking extending for longer than expected by a simple addition of forward and backward masking (Fig. 3.12 right). Surprisingly, varying the position of the stimulus in the hole, i.e. closer to the forward or the backward mask had minimal effects on identification performance.

These findings indicate that studying paradigms combining forward and backward masking, such as rapid serial visual presentation is not just the replication of forward and backward masking alone with a new method, but is actually leading to results that are not predictable based on forward and backward masking alone. Note that the longest effective masking interval in the Uttal (1969a,b) study was around 60ms. This is relatively short, and may be linked to the particular dynamic noise used in the study.

Dichoptic masking, in which the mask is presented to the eye contralateral to the stimulus results in the same general pattern, but with forward masking being much less pronounced (e.g. Turvey 1973), indicating that integration, especially in the retina, is probably responsible for the strong forward masking effect in monoptic masking. This is also supported by evidence showing that in dichoptic backward masking, the energy of the mask plays a minor role as opposed to its prominent role in monoptic masking (Turvey 1973 but see Monahan and Steronko, 1977).

Pattern masking is also affected by the spatial frequency component of the target and the mask. Mitov et al. (1981) used a variation of the study by Green (1981) discussed above. Mitov et al. used gratings as targets and masks in a detection task. The mask was presented for 500ms, the target for 10ms as in the study by Crawford (1947). If both the target and the mask had spatial frequencies below 6c/deg, a strong masking with pronounced transient 'overshoots' were observed. If the mask had a frequency of 6 c/deg, but the target had a higher frequency (18c/deg), much smaller overshoots were observed. And finally, if the mask was high frequency (18c/deg) but the target was low frequency (6c/deg), no transient overshoots at all could be observed. These findings confirm that the magnocellular system, responding best to

lower special frequencies is at the origin of the transient overshoots, and that it exerts the strongest inhibiting influence on itself, as opposed to the parvocellular system.

The similarity between target and mask plays a role in determining the amount of backward masking, with the SOA determining the direction of the effect (Hellige et al., 1979). At short SOA (<32ms), and at all Target:Mask energy ratios, a 'similar' mask (i.e. a mask sharing many features with the target) will produce less backward masking than a 'different' (i.e. a mask sharing few features with the target) mask. At longer intervals (SOA>32ms), the reverse effect is observed: a similar mask produces more masking than a different mask.

3.2.3.3. Para/Metacontrast masking

Para- and metacontrast masking differ profoundly from pattern masking or masking by light in that mask and stimulus are not overlapping in space, but only share some of their contours. The two main stimulus arrangements used in para-/meta-contrast masking is to have a bar as a target, and two bars flanking the position of the target bar as masks, or to have a disk as target and an annulus directly surrounding it as a mask (see Fig. 3.4). The term para- and metacontrast masking were used to refer to cases where the mask precedes the target, and comes after the target respectively. Recently, this use of the term has been abandoned, and replaced by forward and backward metacontrast masking (e.g. Macknik and Livingstone 1999), and this is the way the term will be used in this section.

The phenomenology of metacontrast masking is that at certain SOA, the mask reduces the apparent contrast of the stimulus, rendering it sometimes impossible to detect. Perceived contrast under metacontrast masking follows a type-II masking function: the perceived contrast is high when stimulus and mask are presented simultaneously, then it deceases to a minimum as target and mask are separated by ~60ms, and returns to baseline as target and mask are separated by more and more time.

Metacontrast masking depends critically on the precise alignment of stimulus and mask: as separation between stimulus and mask increases, the masking effect decreases rapidly (e.g. Alpern, 1953; Breitmeyer and Horman 1981). Metacontrast masking is as strong or even stronger under dichoptic conditions (stimulus and masked presented to different eyes), as indicated by Schiller and Smith (1968) and Weisstein (1971), indicating the importance of cortical neuronal processes in

metacontrast masking, although older studies sometimes failed to find dichoptic metacontrast masking (Alpern, 1953), probably due to poor alignment of the dichoptically presented stimulus and mask, which is so critical to metacontrast masking.

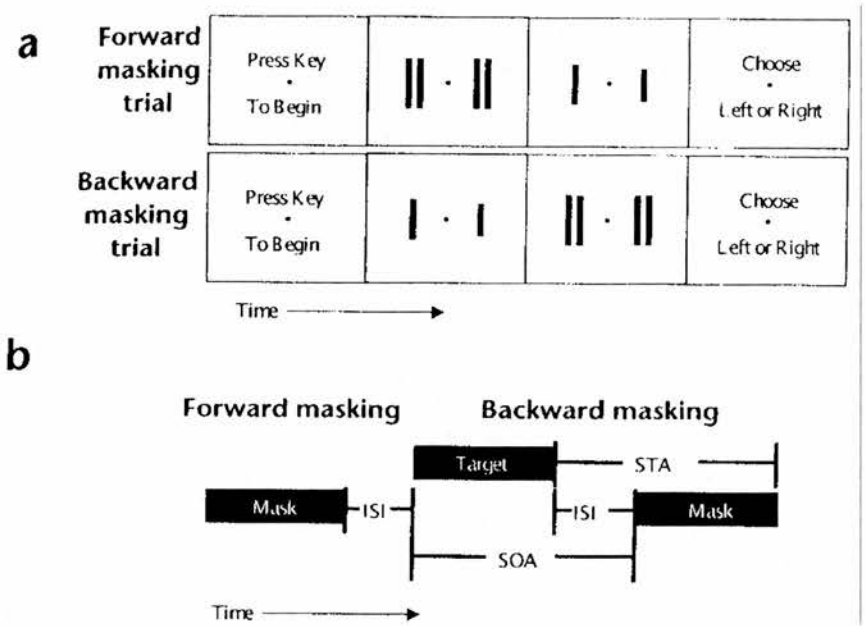


Figure 3.13: Experimental Design (a) the sequence of events in the Macknik and Livingstone (1998) experiment. In backward masking trials the target was presented first, followed by the mask. The reverse happened in the forward masking trials. The subjects task was to identify whether in the target the longer bar was on the right or on the left side. (b) A schematic illustrations of the different parameters that can be used to define the relative timing of target and mask.

Macknik and Livingstone (1999), as a preliminary to their physiological studies, conducted the most extensive study on the influence of stimulus timing in metacontrast. While other studies (e.g. Alpern, 1953; Breitmeyer, 1978; Bridgeman, 1980) varied stimulus duration or ISI separately, Macknik and Livingstone varied both parameters systematically and unravelled the stimulus timing parameters most predictive of metacontrast masking.

Their study measures the performance of human subjects in a visual decision task as a function of target and mask duration. The stimuli used are illustrated in Figure 3.13a. The target stimuli were two bars, one at the left and one at the right of the fixation point. The subject had to decide which of the two was longer. This task was rendered more difficult by the presentation of masking stimuli: a pair of bars

flanking each of the two target stimuli, that were irrelevant to the task. The mask was presented before or after the target. Decision performance was measured as a function of target and mask duration, and the interval between target and mask. They compared a number of different ways to measure the temporal relations between the stimulus and the mask (see fig 3.13b), including the ISI, the SOA and the Stimulus Termination Asynchrony (STA). They plotted the subjects' performance as a function of all three measures, and came to the somewhat surprising conclusion that the stimulus termination asynchrony, was the best predictor for performance in their backward masking task. The worst performance occurred at a STA of approximately 100ms, and performance increasing both at shorter and longer STA. This finding is surprising because one would expect that it is the start of the second stimulus (the mask) that would affect the processing of the first stimulus (the target). But it appears that for metacontrast masking, it is the turning off of the mask that is the most masking event. For the forward masking, they found that it was the ISI (Inter-Stimulus Interval) that was the best predictor for the human performance – a finding consistent with the fact that it is the time of *off*set of the mask that plays the most important role rather than its onset (otherwise, SOA would be the best predictor) – with ISI ~20ms resulting in the worst performance. This finding is not directly transferable to masking by pattern or light. Based on these psychophysical experiments, Macknik and Livingstone designed a very effective visual masking paradigm in which a target bar of 60ms duration is continuously shown in alternation with a 110ms, non-overlapping mask composed of two flanking bars. This arrangement of stimuli results in both strong forward and backward masking of the target by the mask, resulting in almost continuous disappearance of the target from the perception of a human observer. The responses obtained from V1 neurones using this paradigm are reported later.

While metacontrast masking may play a certain role in pattern masking, if parts of the pattern-mask configuration result in directly flanking contours, it is nevertheless not as directly relevant to the thesis at hand, and will not be reviewed in more details from a psychophysical point of view. The interested reader be conferred to for instance Breitmeyer (1984) for an excellent review.

3.2.4. Physiological investigations of masking

A number of physiological studies have investigated brain activity during masking (Table 3.1)

	Masking by light or pattern	Meta-/Paracontrast
Retina	Levick and Zacks (1970) Fehmi et al. (1969)	Bridgeman (1975)
LGN	Schiller (1968) Fehmi et al. (1969)	Schiller (1968) Bridgeman (1975)
V1/Area 17	Fehmi et al. (1969)	Macknik and Livingstone (1998) Bridgeman (1975, 1980)
IT/STSa	Rolls and Tovee (1994a) Rolls et al. (1994b,1999) Kovács et al. (1995)	
Amygdala	Morris et al. (1999)	
Frontal Eye Fields	Thompson and Schall (1999)	

Table 3.1: Investigations of physiological mechanisms of masking as a function of type of masking and brain area investigated.

3.2.4.1. Masking by light or pattern in visual brain areas

The earliest level at which masking has been investigated is the ganglion cells of the cat's retina. Levick and Sacks (1970) presented two brief (2ms) flashes of light into the receptive field of a ganglion cell, while recording extracellular potentials. They used ISI of 0-80ms, and observed for ISI of 80 and 64ms, that the two flashes produced entirely separate responses, with the response to the second flash being weaker that the response to the first (forward masking). For ISI of 48ms or shorter, the two responses start to overlap (=integration), being apparently fully fused at ISI of 16ms or less. The second flash never interrupted the response to the first stimulus, hence, at ganglion cell level, integration but no interruption was observed.

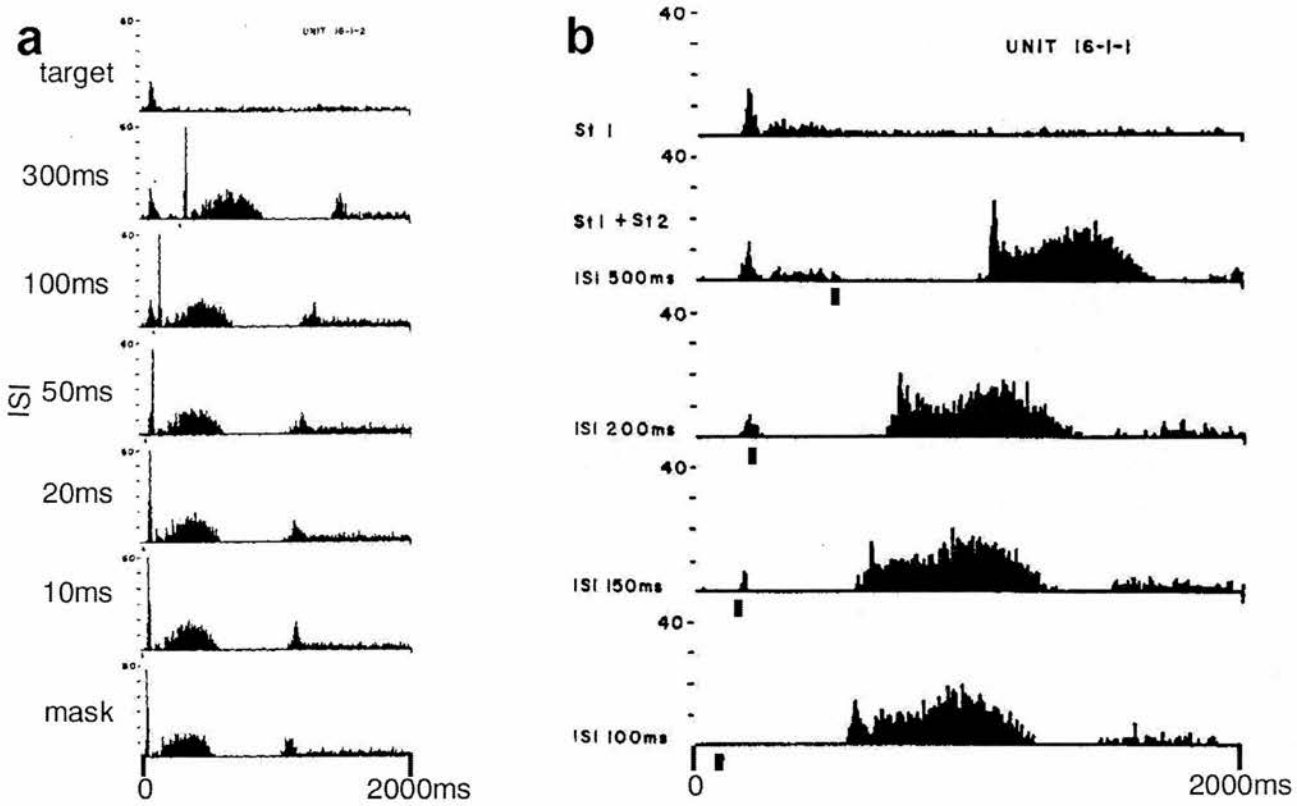


Figure 3.14: Single cell investigation of masking in the LGN of the cat. (a) an on-centre cell stimulated by a dim target and a bright mask (all stimuli were 10ms long). The target is presented at t=0, the mask at the time indicated by the little tick mark under each trace, with the ISI indicated at the left of each graph. As can be seen, the response to the target alone (top line) virtually disappears at ISI of 50ms or less, and the resulting response is almost identical to that to the mask alone (bottom trace). (b) Response of an off-centre cell to the same stimuli. Top trace: response to the target (St 1) alone. Second trace, response to target and mask at an ISI of 500ms, etc. Note the long inhibition caused by the mask. This inhibition curtails or prevents the off-response to the target for shorter ISI [Adapted from Schiller (1968)].

Schiller (1968) investigated backward masking in the LGN of cats. When using a strong mask flash and a weak stimulus flash, he observed that for ISI of 100ms or less, the larger response to the mask will start to overlap with the smaller response to the stimulus (Fig. 3.14a). Due to saturation, the already large response to the mask could not increase any further to summate with the response to the stimulus, resulting at ISI=0 to a response identical to that to the mask alone. This may be a good model for Type-A masking, with an energy ratio in favour of the mask. Using identical flashes for stimulus and mask, he did observe that as the two stimuli move closer to each other, the responses overlap, creating a summed response smaller than the sum of the parts, but larger than a single stimulus alone. Very interesting results

were found when off-centre units were observed rather than on-centre responses (Fig. 3.14b). He used a bright 10ms mask and a dim 10ms target, as for the on-units earlier, and observed for the strong mask an inhibition lasting for about 500ms after mask onset, followed by a strong response occurring ~500ms after the mask. For the weaker target stimulus presented at t=0, the response was weaker and started about 150ms after stimulus onset. Using both flashes, he observed for ISI<500ms that the inhibition created by the mask would curtail the response to the target in off-centre cells. At ISI<150ms, the inhibition caused by the mask literally annihilated the response to the target in the off-centre cells. Hence, while in on-centre cells, the representations of the stimuli summate, leaving some evidence for the presence of the target, in off-centre cells, the inhibition caused by the mask can annihilate all responses to the target. Hence, if one decodes the signal of a pair of on- and off-centre cells, at appropriate intervals, two flashes cause two responses in on centre cells, but only one in off-centre cells. This could all too easily be interpreted as the response to a single longer stimulus, since there is no second off-centre cell response to indicate the termination of the first stimulus stating: "hey, the two on-responses are not one long activity to one long stimulus but two different ones!"

Since the duration of the inhibition in off-centre cells is proportional to the stimulus intensity (e.g. 500ms at 3 log unit luminance above threshold, and 150ms at 0.3 log luminance above threshold), the maximal ISI at which this disappearance of the target off-response is reduced if target and mask have the same energy. Nevertheless, the same effect will occur. This may contribute to integration of stimulus and mask: the off-centre response may be a central cue for the nervous system in detecting when one stimulus starts and another ends.

Fehmi, Adkins and Lindsley (1968) placed their awake monkeys in front of two keys on which a square and a triangle could be back-projected for 10ms (T). The monkey had to press the key on which the square appeared. A bright 20 μs flash (B) was also flashed on the keys at variable ISI, but was irrelevant to the task. The performance of the monkey was perfect for ISI above 30ms and was at chance level at ISI≤15ms. Simultaneously, they recorded evoked potentials from the optic nerve, LGN and visual cortex (Fig. 3.15). In all three sites, clear masking effects were visible. They examined how much of the response to the target (T) remained visible in the joint target + mask (TB) trials by subtracting the response to the mask alone from

the response to the target + mask (TB-B). This TB-B difference curve became increasingly flatter, as the ISI was reduced. Interestingly, at ISI=30ms, where the monkeys performance was perfect, the TB-B curve resembled the T curve for only the first 60ms, indicating that only the first 60ms of the response are needed for identification. Response duration remained around 60ms even at shorter ISI, but in those cases, the amplitude of the TB-B curve became increasingly reduced.

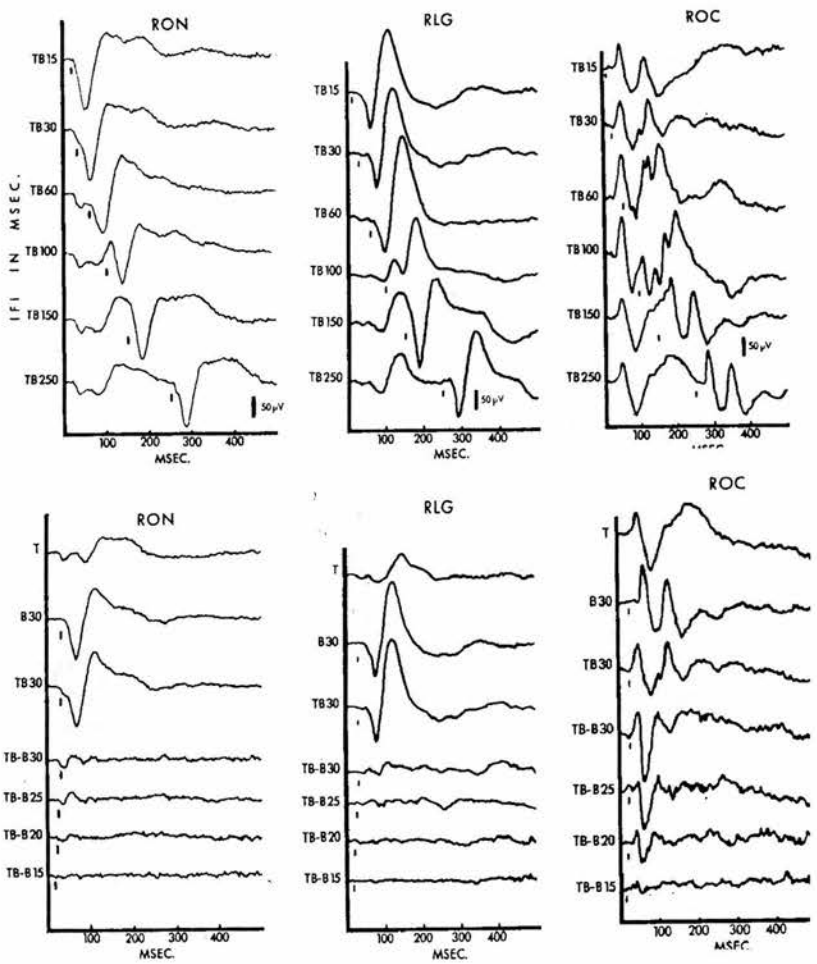


Figure 3.15: Evoked potentials recorded by Fehmi et al. (1968). RON=right optic nerve; RLG=right lateral geniculate; ROC=right occipital cortex. The responses are shown for cases where the target alone was presented (T), the bright flash (B) alone or offset by 30ms (B30), together with cases where both target and flash were presented with an SOA of *t* ms (TB*t*, e.g. TB30). In the bottom row, the response differences between TB and B cases are shown for different ISI, to show how much of the target response remains if the response to the mask alone has been substracted. The black tick under the traces represents occurrences of the mask while the onset of the target occured at t=0ms. Note, that as ISI is reduced from 30 to 15ms and discrimination performance goes from 100% to chance, the duration of the response to the target (i.e. TB-B) decreases to about 60ms, but no less, while the amplitude of TB-B decreased throughout, making it hard to see how long the response really lasts. It is also interesting to see how the amplitude of the evoked potential clearly reflects the

differences in brightness in the RON (i.e. response to target<<mask), but not in the ROC (response to target ≈ mask).

Rolls and co-workers and Kovács and co-workers investigated masking by pattern in single cells of the temporal cortex of the macaque. Rolls et al. (1994a) recorded from single cells in IT and STS, while presenting black and white images of faces for 16ms followed after a variable SOA by a mask composed of an other face or of N and O letters (see Fig 3.16e for stimulus and mask examples). In both cases, if mask and pattern were presented simultaneously a human observer was unable to identify the face, while at SOA=16ms, identification was above chance, implying a Type-I masking.

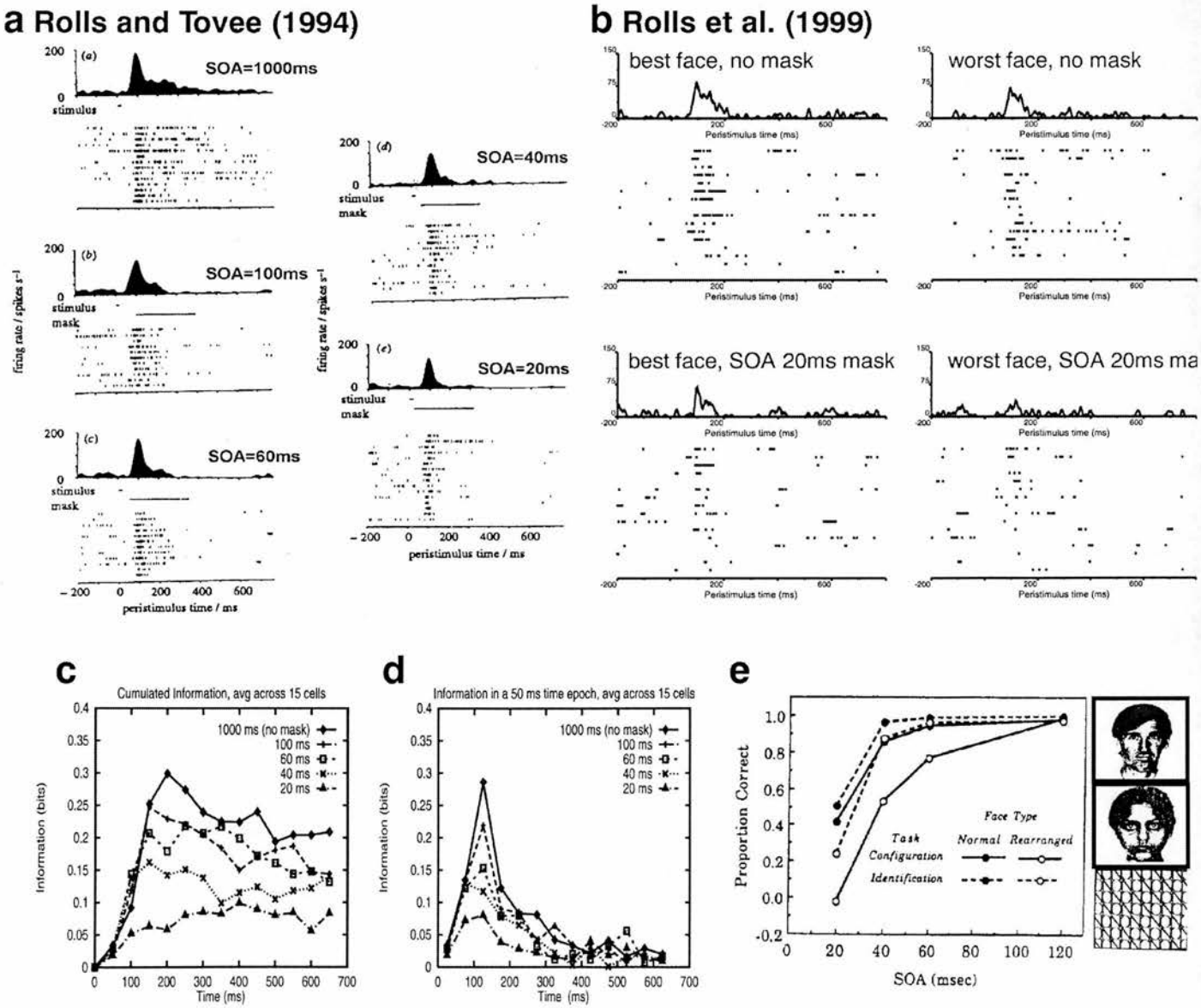


Figure 3.16: Main results from the Rolls et al. (1994a,b and 1999). (a) Spike density function and rastergram at different SOA to the best face image for that cell. Note how the response is increasingly shortened as SOA decreases, but never falls bellow ~60ms even at SOA=20ms [Adapted from and Tovee (1994a)]. (b) Responses to the best and worst face with and without mask. Note how long the response is to the best face with an SOA=20ms mask [Adapted from Rolls et al. (1999)]. (c) Average information contained in spike counts taken from 0 to x ms (as indicated on the abscissa) after stimulus onset. Note how long the cumulative information continues to raise even at SOA=20ms. (d) Average cumulative information contained in a 50ms spike count as a function of the time relative to stimulus onset at which the spike count is centred upon. Note that the information at SOA=20ms is present in an interval ranging from 50 to 150ms post stimulus onset. Note also that the absolute information contained in the response to a SOA=20ms is much lower, even in the first 50-100ms than that to a longer SOA. [adapted from Rolls et al. 1999]. (e) performance of human observers together with example stimuli and masks. In addition to the 6 normal faces, rearranged (i.e. jumbled) faces were also presented. Subject had both to name to whom the face belonged, and whether the face was jumbled or not. For the physiological experiments in which only normal faces were used, only the "identification normal" line is directly relevant. The scores are corrected for guessing, with 0% indicating chance performance [adapted from Rolls et al. 1994b].

Their single cell recordings clearly reflected the influence of the masking (Fig. 3.16). As SOA was reduced, so was the duration of the neural response. Yet at all SOA, response duration was longer than stimulus duration, and for SOA=20ms, the response duration was no shorter than 60ms. An information analysis of their findings (Rolls et al. 1999) indicates, that a backward mask reduces the amount of information about the stimulus contained in the spike counts, but the duration over which the spike counts contain information about the target stimulus are surprisingly not very different in the different conditions (Fig 3.16c,d). Human observers performed largely above chance at about 50% (after correction for guessing, where 0% reflects chance) in a naming task for the same stimuli at an SOA of 20ms (Fig 3.16e, Rolls et al., 1994b), indicating that masking did occur under those circumstances, but that the mask was far from complete. The shortening of the response through the mask indicates an interruptive mechanism of masking. The fact that the shortest responses were no shorter than about 60ms indicates a limit to the interruptive mechanism, and will be discussed in more depth in a subsequent section. The authors themselves favour the idea of a lateral inhibition between adjacent cortical cells to be responsible for this interruption.

Kovács et al. (1995) performed a similar experiment. They used simpler stimuli, namely line drawings (Fig. 3.17 inset), and used the combination of these line-drawings as a mask, thereby ensuring perfect spatial masking of all their stimuli. They first trained two monkeys and two humans to report seeing a particular target drawing by pressing a button. This task was performed with target stimuli of different duration with and without masks. Their results are shown in Figure 3.17 and illustrate how similar humans and monkeys perform in masking tasks, confirming that monkeys are very appropriate for investigations of pattern masking, and that they yield quantitatively identical performance at identical SOA in masking trials. Kovács (personal communication) indicated that the lower score in monkeys for 20ms, unmasked stimuli is probably due to insufficient training because it still improved during the experiments.

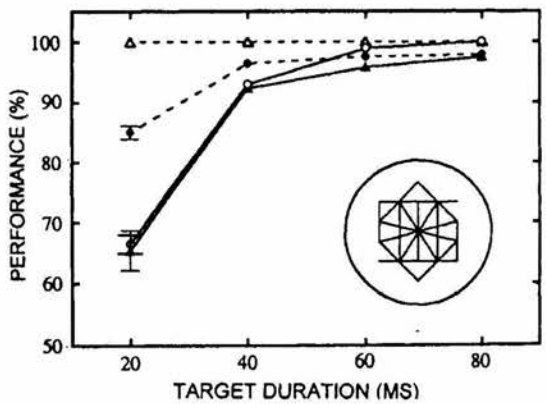


Figure 3.17: Psychophysical results from Kovács et al. (1995). Uncorrected performance is shown as a function of target duration for two human observers (open symbols), and two monkeys (solid symbols). The solid lines represent trials where the stimuli were followed by the mask (inset, composed of the combination of all stimuli) while dotted lines represent the performance for unmasked stimuli.

Using the exact same stimuli, the authors recorded from single cells in IT (see Figure 3.18). As in Rolls et al.'s work, responses indicated that as SOA is reduced, first response duration is affected (160ms to 40ms) but peak firing is not, and then later (40ms vs. 20ms), response duration remains constant around ~60ms but peak firing is reduced. Unlike Rolls and coworkers, they also show, that even without a mask, 20ms and 40ms stimuli produce smaller and shorter responses than longer duration stimuli.

The authors focused on the Best-Worst difference PSTH obtained by subtracting the PSTH to the least effective stimulus ('Worst'=W) from the PSTH

obtained to the most effective stimulus ('Best'=B). This PSTH shows a much shorter response in the masked (M) than in the unmasked (UM) condition. Accoring to this difference curve (M:B-W), a 20ms masked stimulus therefore only produced a 60ms discriminating response before being interrupted by the mask, giving strong evidence for an interruption theory of masking. Unfortunately this line of thinking is flawed. The authors used the integration of all stimuli as their mask. By definition, this makes it impossible to differentiate integration and interruption theories of masking. Interruption theory predicts that after occurrence of the mask, the neural representation of the target should be interrupted and replaces by that of the mask. Integration theory indicates that the response to the target should be replaced by a response to the integration of target and mask. Unfortunately, the integration of target and mask is the same as the mask in this case.

In their experiments the cells responded strongly to the mask (comparing UM:W with M:W in Fig 3.18). This mask response was on average so strong, that it is indistinguishable from the sustained sections of the response to the best stimulus alone (UM:B). Hence, the fact that, in the masked condition, 60ms after response onset, the response to the B and W stimulus is identical (i.e. M:B-W=0), can be equally attributed (i) to the cell continuing to responds to the best stimulus in a sustained fashion, or (ii) to a sudden transition to represent the mask.

In their Figure 5, they illustrate a cell for which the mask was not effective, but in that cell masking did not result in a shortening of the response, but rather in a complete annihilation of the response, making it impossible to judge its duration.

Hence their otherwise excellent study shows interruption at long SOA, (comparing B responses in masked and unmasked conditions in Fig.3.18), but fails to differentiate between integration and interruption at short SOA due to the choice of their mask and the strong single cell response to the mask.

Both the studies by Rolls and co-workers and Kovács and co-worker show, that at the level of the temporal cortex, the spike counts in single cells correspond to some extent with the psychophysically measured target perception. Rolls et al. showed that reducing SOA produces a decrease in mutual information at the level of single cells in STS (Rolls et al., 1999), and a reduction in human identification performance (Rolls et al., 1994b). Kovács et al. (1995) show also a decrease in both human and monkey psychophysical performance, and state that a ROC analysis of the

spike counts occurring in the first 80ms of the responses was related to psychophysical performance.

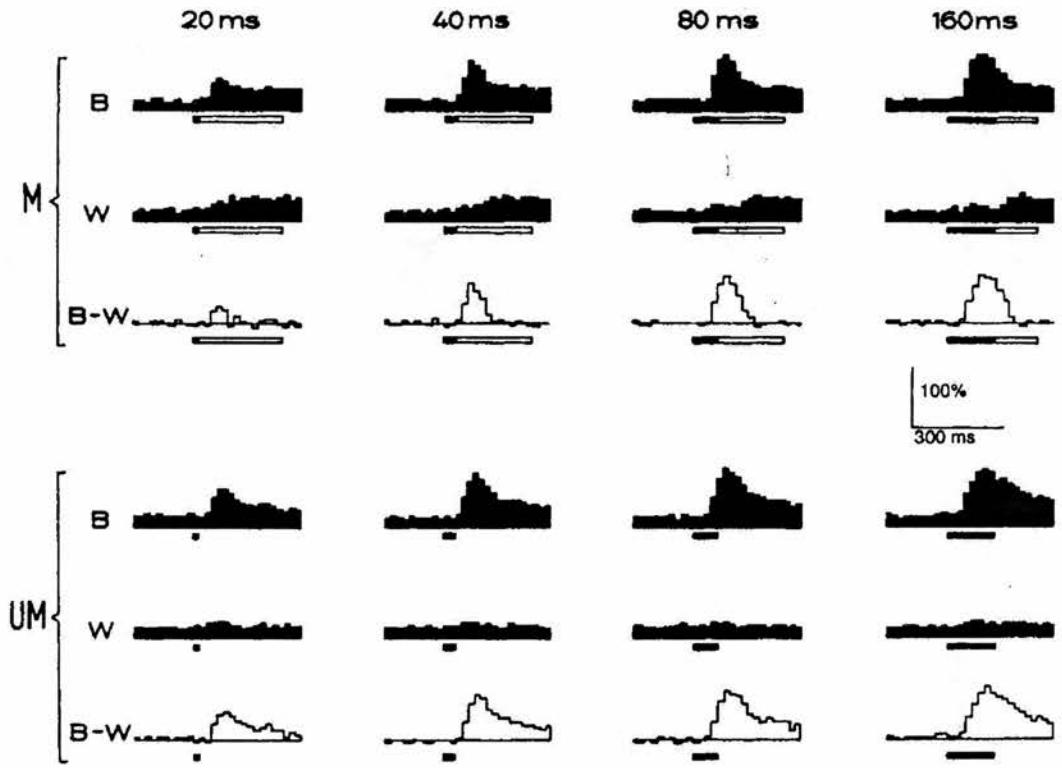


Figure 3.18: Averaged normalised PSTHs of 54 shape selective IT neurones at target durations of 20 to 160ms. Target and Mask presentation are indicated by solid and open bars respectively. Rows 3 and 6 represent the difference between presentation of the best (B) and worst (W) in the Masked (M) and unmasked (UM) condition respectively. Notice the relatively strong response to the mask apparent when comparing responses to the worst stimulus in the masked and unmasked condition. [Adapted from Kovács et al., 1995].

Physiological conclusions regarding pattern masking in the visual system: Altogether, it thus appears, that at the various stages of the visual system, firing patterns are affected by masking. It appears that for SOA between about 60 and 500, an interruption theory of masking is successful at explaining the results: peak firing rate is unaffected while response duration is. At SOAs progressively shorter than 60ms, response duration does not decrease below about 60ms, but peak firing decreases. The exact mechanism involved - interruption or integration or even simple competition – has not been unravelled through these studies. Only in the retina do we have evidence for integration for SOA<60ms but never interruption.

3.2.4.2. Pattern masking in brain areas beyond the visual system proper

What happens beyond the visual system proper is less clear. Single cells in the frontal eye field have been studied by Thompson and Schall (1999). In their task, a monkey had to saccade to a target that was masked. In this way, on a trial-by-trial basis, data was available about whether the target had been processed, and how much the neurone fired. Neurones fired more in 'hits' than on 'miss' trials, and more in 'false alarm' than in 'correct rejection' trials. Nevertheless, neurones did fire in cases in which the monkey did not report seeing the target, indicating that the cells receive information about the stimuli at SOA=16.7ms, although there was little or no reported perception of the stimulus at that SOA. This is different from results in IT, where there has never been a report of a masking condition in which there is stimulus discriminating neural firing to a pattern but no reported perception of that pattern. Yet it has to be said, that no trial-by-trial perceptual report has so far been measured in conjuction with single cell data under masking conditions within IT cortex.

In the amygdala, positron emition tomography indicates activation of the right amygdala to the presentation of a visual stimulus of emotional valence that was not consciously perceived because it was masked. Morris et al. (1999) presented pictures of faces, and associated one of the angry faces with a loud 100db white noise burst. After this fear conditioning, they scanned subjects, while presenting pairs of faces, with the first face being presented for 30ms followed immediately by the second face. Subjects had to report seeing the angry face they had been conditioned to by pressing a button. If this target face was first in the two-face sequence, it was never detected. Nevertheless the right amygdala activation and the skin conductance indicated that the fear conditioned face had been successfully processed. The authors implicate an extra-thalamic pathway, involving the superior coliculus and the pulvinar nucleus in the processing of the masked feared faces. This pathway, avoiding the normal ventral visual pathway may explain why the amygdala but not the inferior temporal cortex seems to be activated in masking without perception. The same pathway may be responsible for blind-sight, and would hence mediate visual processing without awareness, and explain the unconscious galvanic skin response to unseen fearful stimuli.

3.2.4.3. Metacontrast masking in the brain

Schiller (1968) investigated metacontrast masking in the LGN in the same study in which he investigated masking by light. In the LGN he found that flashing an annulus after flashing a disk causes the response to decrease, in accordance with the centre-surround antagonism in the LGN. Yet, as he varied ISI, the response increased monotonically as ISI was increased from 0ms, indicating a type-I masking curve that cannot by itself explain the type-II perceptual effect of metacontrast. Hence metacontrast seems to rely on mechanisms higher than the LGN, as is also indicated by the fact that dichoptic metacontrast masking is as strong as monocular metacontrast masking (Schiller and Smith, 1968; and Weisstein, 1971).

Bridgeman (1975) diverged from the standard metacontrast arrangement by presenting one bar as a stimulus, and just a single bar flanking the other first bar. Under those conditions, what is the stimulus, and what is the mask is irrelevant, forward and backward masking can be examined simultaneously. The bars were presented so that the midline between the two bars was in the centre of the receptive field. He recorded from neurones in the optic tract, the LGN and Area 17 of curarised but awake cats. In the optic tract no physiological metacontrast masking effect was observed. In the LGN, the early component of the firing did not show metacontrast masking effect, while the later components (later than 160ms after stimulus flash) decreased if the second flash occurred within about 80ms of the first, both in on- and off-cells, but following again a type-I masking curve, with maximum masking at ISI=0ms, as in the single cell study of Schiller (1968). In Area 17, two types of cells had to be differentiated. Some cells showed only an early peak (before 160ms), and showed Type-I masking effects with maximum masking at SOA=0. The second type of cells showed a two peaked firing pattern. A first peak (0-120ms) was followed by an inhibitory phase (120-160ms), followed by a second, later peak (160-260ms). This type of cells showed no masking in their first peak (spike counts 0-120ms), but their late peak (measured as spike counts from 160-260ms) showed the type-II masking typical for psychophysical metacontrast masking measures. This metacontrast correlate occurred when the late components of one stimulus fell in the inhibitory period of the other stimulus. This led Bridgeman to argue that the early components of the responses may mediate psychophysical detection of a stimulus (which is not affected by metacontrast under these conditions), while the later components mediate contrast perception (which is affected by metacontrast masking).

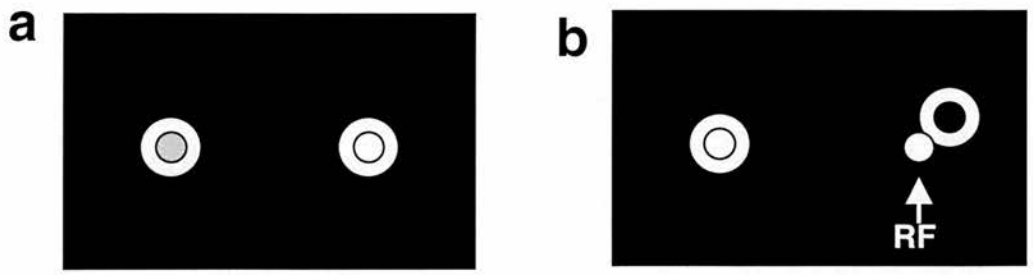


Figure 3.19: Two of the tasks used by Brigeman (1980). In the brighness discrimination task, two identical annuli are shown to both sides together with a 100% brightness disk on one side, and a 80% brightness disk on the other side. The dim side was alternated randomly, and the monkey had to indicate the side on which the brighter of the two stimuli was placed. In the metacontrast trials, (b) the receptive field (RF) of a cell was mapped (here it is on the right side). At time t=0, 3 of the 4 elements are flashed simultaneously: the disk alone on the side of the RF, and the disk and the annulus on the other side ('simultaneous' side). 80 or 120ms later, the annulus is flashed on the RF side ('delay side', schematically, the temporal offset is represented by a special offset of the annulus). Here all 4 components had the same brightness, and the monkey chose the simultaneous side as looking brighter on a majority of trials due to metacontrast masking. The two types of trials were interleaved, to avoid that the RF side was always the dimmer side.

Bridgeman (1980) later investigated metacontrast masking in monkeys performing a brightness discrimination task. He trained the monkeys to report which of two stimuli seemed brighter (see Fig. 3.19). In brightness trials (Fig. 3.19a) the stimuli had a physical difference in brightness. In metacontrast trials (Fig. 3.19b), the stimuli were identical, but the ISI introduced between the annulus and the disk on one side made it appear dimmer. By measuring the effect of varying the ISI on the monkeys report of which of the stimuli seemed brighter, Bridgeman determined a masking curve for the monkeys, that resembles that of humans, and had a minimum apparent brightness at ISI=80ms for one monkey and ISI=120 for the other monkey. He then recorded from V1 while the monkeys continued reporting which of the two stimuli was brighter. During recording he always presented the simultaneous disk+annulus to the side contralateral to the receptive field, and presented the disk followed after the optimal ISI (80 or 120ms depending on the monkey) to the receptive field. He interleaved those trials with brightness trials (Fig. 3.18b) to avoid that the monkey based his decision on criteria other than apparent brightness. For metacontrast trials, he found, that the firing rate in the late part (160-260ms post

stimulus onset) of the response was higher in trials in which the monkey chose the side of the receptive field to be brighter, compared to trials were the contralateral side appeared brighter, indicating a firing rate-apparent contrast correspondence for the late, but not the early components of the response. This led him to attribute the later components to iterative processes: information goes from the LGN to V1 (early components), on to later cortical areas, and is then reverberated back to V1 (late components). The late components can now reflect more cognitive effects, which result from processes higher up in the visual cortex.

Macknik and Livingstone (1998) continued his work and used the "Standing Wave of Invisibitily" paradigm, in which a single target bar is continuously flashed for 60ms in alternation with a 100ms mask composed of two bars flanking the target bar to investigate metacontrast masking in V1. Under these conditions the centre bar disappears from human perception due to metacontrast masking. They recorded from neurones in V1 of awake behaving *Macaca mulata*, comparing the responses of the cells to the masks alone, the stimulus alone and the two combined. 77% of 30 cells, showed inhibition of the transient on response to the stimulus in the target+mask condition compared to target alone condition (Fig. 3.20a).

To investigate the masking effect on the later components of the response, they used a variant of their procedure in which a single Mask-Target-Mask triplet was presented (Fig. 3.20b). Under such conditions, both the transient on and the transient off response were inhibited. To separate the effect of forward and backward masking, they finally presented either a Mask-Target or a Target-Mask pair alone (Fig. 3.20c). The forward masking condition resulted in an inhibition of both the onset and the offset response to the target, whereas the backward masking condition resulted in a selective inhibition of the transient off-response to the stimulus. The fact that psychophysically both forward and backward masking alone reduce the perceived contrast indicates that both the transient onset and the offset components of the responses contribute to the visibility and perceived contrast of a stimulus. In anaesthetised preparations, without eye movement, they could position the stimuli so that the mask produced no response, and could reconfirm these findings (Fig. 3.20d). They studied in addition the effect of masking in long presentations of the stimulus (334ms, Fig. 3.20e) and found that the sustained response in V1 neurones happening between the transient on- and offset components can be inhibited by an mask flashed during the target presentation, but under such conditions no masking is reported by

human observers, indicating that unlike the transient on- and offset response, the sustained response is not necessary to the perception of the stimulus.

Hence, both the transient on- and off response correlate positively with conscious visual perception, or at least with the ability to judge which of two lines is longer, whereas variations of the sustained response between these transient on and off response seem not to correspond to changes in consciousness. Of course the correlation of the transient components do not prove that the transient response is V1 result in consciousness. They *may* very well just be needed to activate subsequent stages which themselves might create the conscious experience.

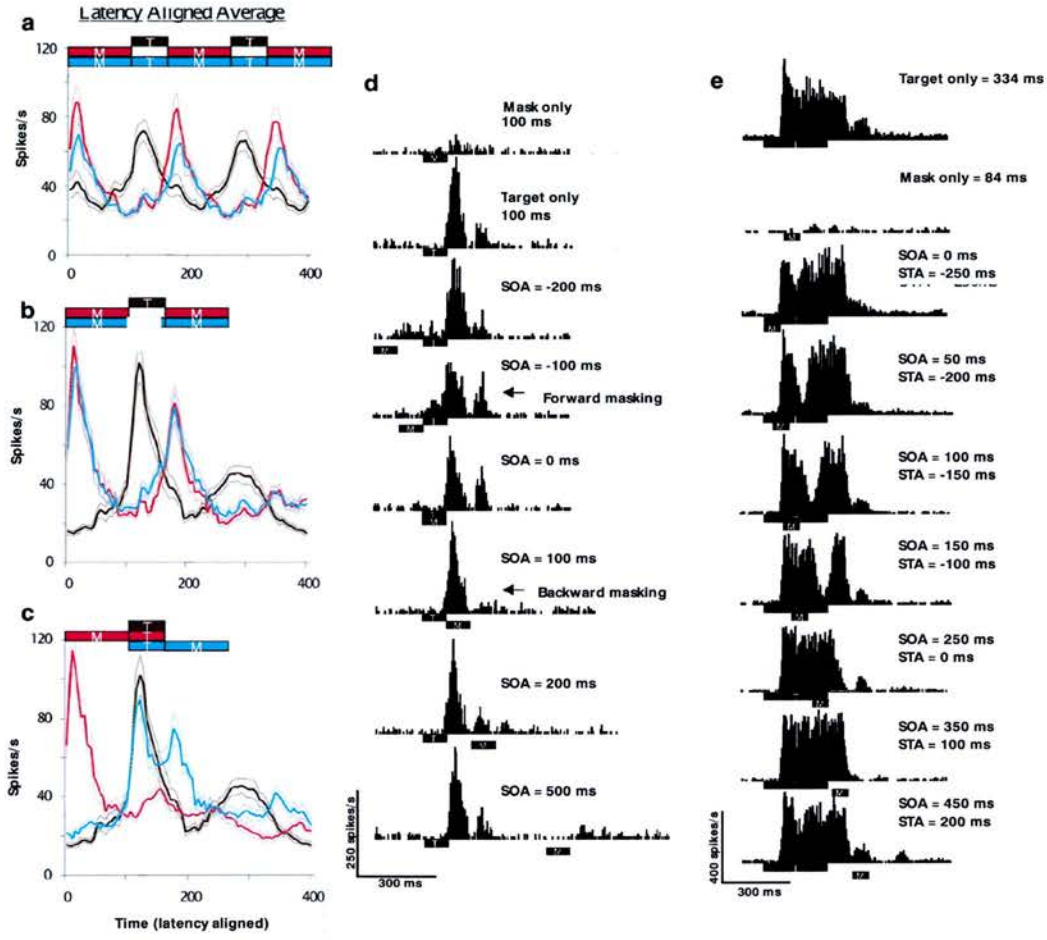


Figure 3.20: findings of Macknik and Livingstone (1998). (a-c): latency aligned responses of 30 neurones in V1 of the alert rhesus monkey. (a) Response to the target (T) alone (i.e. the single bar), the mask (M) alone (the two bars) and both in alternation. Note how the response to the target almost disappears. (b) Same as in a, but for a single M-T-M triplet, note how both the transient on- and off-responses disappear. (c) Target alone and forward and backward masking. Note how forward masking disrupts both the early and the late peak of the target response, while backward masking only disrupts the late response, as found by Bridgeman

(1975,1980). (d) Multiunit recording in the anaesthetised rhesus. The anaesthetised preparation enabled a more precise placing of the target onto the receptive field resulting in no response to the mask alone. The recording clearly shows the type-II masking function, with forward masking effecting the early peak and backward masking affecting the late peak (or after-discharge). (e) Multiunit recording to a long target presentation (334ms) in the squirrel monkey. Note how a mask can inhibit the transient on- and off-response as well as the sustained components. Yet in humans, perceptual masking is only reported in conditions that here produced transient on- or off-response disruptions, but not in those producing sustained component inhibition.

In an elegant optical imaging study, Macknik and Haglund (1999) confirmed the effect of metacontrast masking on V1 by showing that the activation pattern associated with the target in their "standing wave of invisibility" disappears from the cortex when the target is flickered in alternation with the masking double bar.

Conclusions of metacontrast physiology: Metacontrast masking is a much more elusive phenomenon than pattern masking or masking by light. Psychophysically it is characterised by a type-II perception (see Fig. 3.6b for Type-I vs Type-II), where little masking occurs at SOA=0ms. Under the same conditions single cells in the retina and in the LGN fail to show type-II masking: all show the strongest masking at SOA=0ms, in accordance with the classical centre surround antagonism of these cells. Only in V1 has it been possible for Bridgeman and his student, Macknik and co-workers to identify neural correlates of type-II metacontrast masking, a finding compatible with the psychophysical finding, that metacontrast is just as strong for dichoptic stimuli. These V1 studies revealed the importance of the late components of the response for backward metacontrast masking (called afterdischarge by Macknik and Livingstone), and of the early, transient components in forward metacontrast masking. If type-II masking is apparent in V1, one would expect it to be carried forward to later cortical areas, such as V4. Surprisingly, a recent paper by Kondo and Komatsu (2000) using a disk as a stimulus and an annulus as a mask failed to find type-II effects under metacontrast masking in V4: they found type-I metacontrast masking with a maximum at SOA=0ms. It is unclear why type-II masking is found in V1, in accordance with brightness perception and psychophysical discrimination performance, while in V4, closer to the areas often associated with perception, type-I metacontrast masking is found, as in the LGN and the retina.

3.2.4.4. Integration, interruption and minimal response durations: lessons from physiology

At a behavioural level, it is hard to distinguish interruption from integration. Physiology can give us some answers. Interruption in backward masking would be defined as a neuronal response selective for the target that is long in the unmasked condition, and shorter in the masked condition. More particularly, the duration of the response should be directly related to the SOA. The general finding in the literature is that this is the case, but that there is a lower limit to this process. For SOA larger than ~60ms, response duration seems to reflect SOA (Schiller 1968, Fehmi et al 1968, Kovács et al. 1995, Rolls and Tovee 1994a; Rolls et al. 1994b, 1999). For SOA decreasing below 60ms, the response duration seems to remain constant around ~60ms, but the amplitude of the response (measured in terms of information content, receiver operator curve (ROC) or peak firing rate) decreases (Schiller, 1968; Fehmi et al., 1968; Kovács et al., 1995; Rolls and Tovee, 1994a; Rolls et al. 1994b, 1999).

This value of about 60ms is similar to the minimal response duration found in the retinal ganglion cells to very brief (2ms) flashes of light, and is also similar to the maximum duration in which Bloch's law (Bloch, 1885) is valid. Bloch's law states that brightness perception within an interval of ~60ms is directly proportional to the stimulus energy given by the product of intensity and duration of the flash. Hence making a flash twice as long, or twice as bright results in the same apparent brightness change, but only up to a duration of about 60ms, indicating the integrative nature of the visual system for intervals up to about 60ms.

Together this seem to suggest, that interruption proper occurs for SOA>60ms, causing the duration of responses to be terminated by the occurrence of the mask, while some form of integration is prevailing at shorter SOA, leaving the response duration at 60ms but reducing the amplitude of the response, possibly by reducing the apparent contrast of the target by integration with the mask, just like the contrast of the stars in the sky is reduced by integration with the scattered sunlight during the day.

Both Kovács et al. (1995) and Rolls et al. (1994, 1999) favour the interruption theory of masking, and state that at 20ms SOA responses are no longer than the stimulus. Yet their own figures tell us otherwise (see Figure 3.21). Rolls et al.'s spike density functions in the SOA=20ms is clearly elevated for more than 20ms, moreover despite the abovementioned difficulties in interpreting the Best-Worst curve in the

Kovács et al's investigation, their Best-Worst curve is elevated for 60ms for SOA=20ms.

One may counter, that these longer average responses are due to short individual responses but jittered in time from trial to trial. There are two responses to that statement. First, Levick and Sacks (1970) showed that this is not the case in the retina. They examined single trials, and observed response durations of no less than about 60ms even within single trials. If the same is done in Fig. 3.21, many responses in the 20ms SOA condition much outlast stimulus duration. Second, and more fundamentally, each single cell receives the input of not one but many cells. Hebb's law will encourage input from pre-synaptic neurones with similar tunings (since neurones with similar tuning will tend to fire simultaneously, and hence reinforce their synaptic connection in the development of the visual system). Hence, in a single trial, a postsynaptic cell will summate many redundant synaptic inputs. This process is not unlike averaging many trials from a single cell. Hence examining the duration of spike density functions, in which many trials from one cell are averaged, is in a way a model of what a cell does, averaging in a single trial the input from many cells. Unfortunately we do at present not know if the small latency jitter observable between trials of a single cell (e.g. Fig 3.21b) will be correlated with the jitter in other cells. That is to say: if on a trial one cell fires a bit later, do all the other cells tend to fire a bit later as well? Until we have a response to that question, it is impossible to know if pooling many cells in one trial would give a more narrow response to a brief stimulus than pooling many trials of the same cell.

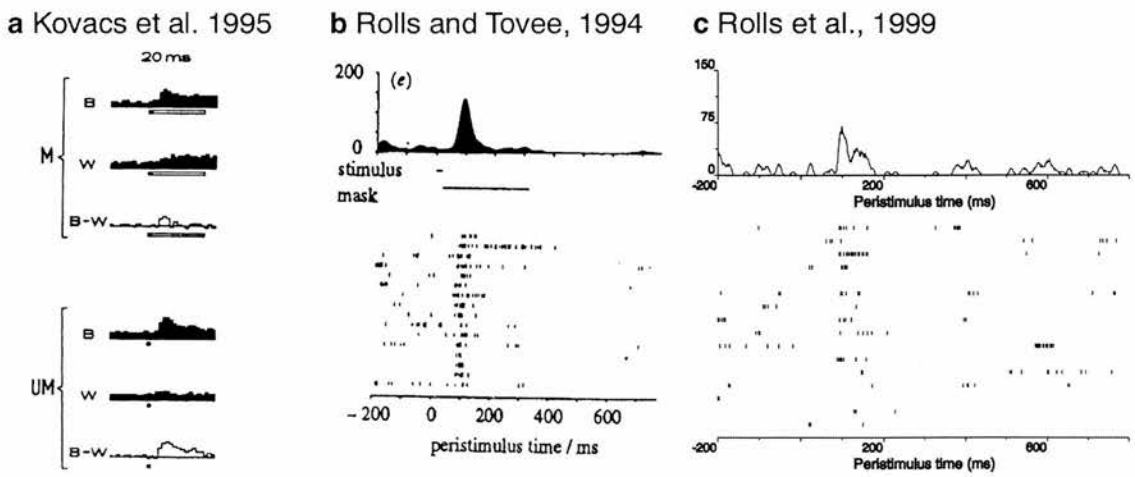


Figure 3.21: Response durations in the temporal cortex in masking with SOA of 20ms. (a) Adapted from Kovács et al. (1995), this represents the pooled normalised PSTH of 54 shape selective neurones in IT. Conventions as in Fig. 3.18. Note the high response in the masked W versus unmasked W indicating the high response caused by

the mask, which subtracted from the B will create the appearance of a very brief response. (b) Spike density function and rastergram for the presentation of a 16ms face followed by a mask at SOA of 20ms. Many individual responses clearly outlast stimulus duration, as is reflected in the rather long lasting response in the spike density function. [Adapted from Rolls and Tovee 1994a). (c) same convention as (b), but adapted from the 1999 paper, showing a different cell.

Generally it is hard to pin down the neural locus of integration or interruption. The psychophysical literature indicates that integration is likely to occur at pre-cortical loci while interruption has a more cortical locus. This is due primarily to observing that masking forms that occur only if the target and the mask are presented to the same eye, depend on the relative energies of target and mask, and thus resemble the double exposition of a photo plate, while those forms of masking that also work if TS and MS are presented to contralateral eyes are relatively immune against energy modifications, which indicates a more interruptive nature. From a more physiological point of view, this distinction is weak, for at the cortical level, the response to a stimulus is relatively unaffected by stimulus energy (e.g. Fehli et al., 1969). Hence, integration of two stimuli at cortical level should not be energy dependent. Therefore, integration of features could also occur at cortical level, and would not be affected by stimulus energy. This is well documented by the evoked response potential (ERP) study of Fehli et al. (1969, our Fig. 3.15). Their stimulus had a lower energy than their mask. The ERP produced by the two stimuli are very different in amplitude in the optic nerve and the LGN, but almost identical in V1. In addition, they observe interruption of the evoked response to the stimulus even at pre-cortical levels. Hence, it is likely that both integration and interruption occurs to some extend at all levels of the visual system, with the exception of the photoreceptors, in which only integration occurs.

3.3. Persistence and Masking: Conclusions

The review of persistence and masking demonstrate a little war in the brain. One party is persistence. It is the tendency of the brain to keep the representation of a stimulus alive for about 60-200ms after the stimulus disappeared. Persistence is beneficial under many circumstances, because it allows the visual system to continue processing a brief stimulus after it has disappeared, thereby freeing the brain from some of the time constraints of the outer world. The internal state of the brain is then different

from the external state of the world. In the lab, this is beneficial in partial report paradigms, but, much closer to everyday life, each time we blink with our eyes, the outer world briefly disappears, and yet, our brain is well advised to continue processing whatever was on the retina just before we started closing our eye lids.

On the other side, there is a cost to persistence. People remembering old green computer-screens, on which each moving point was followed by a trailing ghost, are aware of the blurring consequences of persistence in a changing world. Movements would look as if taken with a photographic camera at an exposure time of about 60ms: a rather blurry picture. Without saccadic suppression, each saccade would result in the integration of the image before and after the saccade. This introduces the second party on the battlefield: masking. Masking is akin to binocular rivalry: it ensures that no two mutually exclusive interpretations of the visual input are represented in the brain at the same time. Masking comes in two flavours: in backward masking a new stimulus terminates the persistence of an old stimulus thereby preventing the old stimulus to blur the new picture. In forward masking, the old stimulus inhibits the representation of the new stimulus, resulting in the same prevention of double representation. From the physiology, we know that the transients of images are important. Under normal viewing, when the new object appears, the transient on-response will place it at an advantage in the masking fight against the old stimulus, of which only a sustained response was left. This should give the new stimulus a systematic bias, resulting reliably in the new stimulus winning. Hence, this battle between persistence and masking ensures that if nothing much new happens, past stimuli will continue to be processed. If something new does happen, it will usually be an effective mask, and stop the persistence of the old stimulus: letting bygones be bygones. Hence this war fought out between the neural representation of stimuli in masking ensures that the best of both worlds are combined: a crisp image without much blurring, and at the same time maximum processing time if nothing new happens.

The limits of the system show up at SOA below 60ms. For some reason, the normal interruption of persistence through the old stimulus seems to fail. Rather than having a period where only the old stimulus is represented followed by a new period in which only the new one is represented, with a brief transition period (not unlike binocular rivalry), at those short SOA, the representation of the old stimulus is weakened by the presentation of the new stimulus. Both are co-represented to some

extent, and the brain is less able to process either one. Perceptual integration of the two stimuli may occur, rendering the weaker one invisible.

4. Stimulus competition in the visual system: III. A unifying perspective

Chapter 2 reviewed the literature on binocular rivalry. It demonstrated, that early in the visual system (V1/V2), both rivalling neural interpretations of the dichoptic visual input are represented. It is only later in the ventral stream, in area IT and STS, that a single interpretation emerges as a winner from what appears to be a competition between the neural representations of the two interpretations of the visual input. Indeed, strong rivalry between interpretations of the visual input is not restricted to cases in which the competing stimuli enter through different eyes: sometimes, two pattern present in the same image can compete against each other, using mechanisms that have been shown to be linked to those of binocular rivalry (Andrews and Purves, 1997). Also, even in binocular rivalry, competition occurs between the interpretation of the stimuli and not between ocular channels: sometimes the percept resulting from binocular rivalry is not one of the two ocular channels, but rather a meaningful composite created by combining some features from one eye, and some features from the other eye (Logothetis et al., 1996; Kovács et al., 1996). Hence, it appears, that the neural mechanisms responsible for binocular rivalry are not restricted to the case of dichoptic stimulation.

Chapter 3 reviewed the literature on persistence and masking. The view emerging from this review, is that the visual system internal state differs from the external state of the world: the visual system displays a 'ballistic' tendency to continue to represent a stimulus after it has disappeared from the outside world. This represents an automatic, very short-term memory for the stimulus, which unlike other forms of memory do not depend upon volitional mnemonic processes. A new stimulus presented after the first will act to terminate the ballistic persistence of the first stimulus. This process can be seen as the representation of the second stimulus competing against the representation of the first. Indeed, forward masking, in which a stimulus presented earlier reduces the representation of the later stimulus indicates that masking is not a one way street, in which the newer stimulus masks the older stimulus: it is more a mutual competition between the stimuli.

Reviewing both literatures, it becomes evident, that they are similar: in both cases, the neural representations of mutually exclusive interpretations of the visual input compete against each other. In one case, the competition is between interpretations of two simultaneously presented stimuli, one presented to each eye, and could be called 'binocular competition'. In the second case, the competition occurs between two stimuli that are presented sequentially and not simultaneously, and could be called 'temporal competition'. But are those two situations really different?

As mentioned earlier, binocular rivalry seems not to be so much the competition of two eyes against each other than the competition between two interpretations of the dichoptic visual input. The eye of origin appears to be a supplementary factor, that makes competition more pronounced (Kovács et al., 1996), but not a necessary factor, and binocular rivalry appears to be physiologically linked to monocular rivalry (Andrews and Purves, 1997). Hence, what appears to be crucial to binocular rivalry is not the dichoptic input, but the existence of the representation of two mutually exclusive interpretations of the visual input in early visual cortex.

In masking, the competition occurs between stimuli that are not present at the same time in the outside world. That appears to make masking quite different from binocular rivalry. Yet, as demonstrated in Chapter 3, due to neural persistence, stimuli that do not overlap in time in the outside world will have neural representations, which do overlap in time. In addition, because of the differences in response onset latencies both within and between areas, cells with earlier latencies will already respond to the mask while cells with later latencies still respond to the target, creating even more temporal overlap between the representations of stimuli that did not overlap in time in the outside world. Hence, in masking as in binocular rivalry, stimulus representations of mutually exclusive interpretations of the visual input will exist at the same time in the brain. Although the stimulus configuration in the outside world is quite different in the two cases, the configuration of stimulus representations in the brain is quite similar.

There is a further similarity between binocular rivalry and masking. Throughout Chapter 3, it became clear, that the stimulus competition occurring during masking is not an immediate process: the presentation of the mask does not immediately terminate the neural representation of the preceding stimulus. Instead,

the process occurs gradually over a period of ~60ms. In a personal communication, Sheinberg confirmed that something similar occurs in flash suppression (see Chapter 2, Figure 2.3): responses in flash suppression are delayed by about 60ms compared to responses produced by the same optimal stimulus in the same cells if no other stimulus was presented before. This demonstrates that it also takes about 60ms, for a new stimulus to compete and win against the old stimulus in flash suppression, a special case of binocular rivalry.

In this Chapter, I would like therefore to argue, that binocular rivalry and masking are two expressions of a same underlying mechanism: 'stimulus competition' or more precisely, the competition between the neural representations of mutually exclusive interpretations of the visual input. To plan actions, our brain needs to settle on a single interpretations of the outside world. While often this is an easy task, in other, more ambiguous situations, it is not. If a monkey is shown an attractive female through one eye, and a leopard through the other eye, he has to decide to run away or seek closer acquaintance with the female. If the two stimuli are such that they cannot exist at the same time, because they occupy the same space, the brain appears to resolve the incongruity by interpreting the visual input as one of the two stimuli, suppressing the other one interpretation for a while. It does so in an 'intelligent way' combining if necessary elements from different eyes to make sense of the stimuli (Kovács et al., 1996). If such competition would not occur, the monkey would perceive the leopard and the female at the same location in space and planning a course of action would be extremely difficult, potentially costing him his life. While suppressing a single percept can lead to mistakes (suppressing the leopard interpretation, if it really was a leopard can be fatal), perception fluctuates between the alternatives, reducing the probability of fully misinterpreting a stimulus (Leopold and Logothetis, 1999).

And old story makes the point quite nicely: A thirsty donkey realises, after a desperately long walk, that he is standing exactly half way between two ponds. The donkey thinks, and thinks: "which pond should I go to?". Thinking to long, he dies of thirst, just because he could not decide what pond was closest. Action is about taking decision, and taking decisions is about excluding alternative possibilities.

Situations of binocular rivalry resemble the story of the donkey: two views of the world are equally likely to be true, yet they are mutually exclusive. To plan a course of action, one of the two alternatives need to be inhibited for action to be

initiated. The stimulus competing mechanisms responsible for binocular rivalry may serve this purpose: stimuli compete against each other, to produce a single congruent representation of the world - probably in the service of planning a course of action.

Masking serves the same purpose: avoiding that two mutually exclusive stimuli are perceived at the same time in the same place. Indeed, it appears, that each time we make a saccade, masking ensures that the view of the world before the saccade does not persist into the view of the world after the saccade, and in addition, that the blurry perspective of the world during the saccade is masked by both the preceding and the following clear views during fixations (Volkmann, 1986; Judge et al., 1980). Masking, and not an efference copy of the oculomotor command or the proprioceptive feedback from the eye appear to be the prime causes responsible for avoiding saccades bluring our perception (Volkmann, 1986; Wurtz, 1969a,b; Judge et al., 1980). The adaptive value of such masking is obvious and twofold. First it avoids blurry perception during a saccade. Second it avoids confusions between what is part of the new fixation and what was part of the old fixation. If a monkey sees a leopard on his right, and an escape route on his left, he will saccade from right to left and then run to the left. Masking ensures, that when he looks to the right, he just sees the leopard, and when he looks to the left, he just sees the escape route. Without masking, due to neural persistence, he would perceive a leopard on his escape route on both sides of the world, and would be unable to plan a course of action.

These simple examples demonstrate how stimulus competition or more exactly the competition between mutually exclusive neural representations of the visual input is necessary to construct a unified percept of the world. This unified percept of the world facilitates the planning of a course of actions. Stimulus competition also occurs in a number of other perceptual situations, such as bi-stable visual illusions (Fig. 4.1), where it becomes very clear that the competition occurs between mutually exclusive interpretations of the visual input and not necessarily between two separate visual stimuli.

Stimulus competition seems to occur whenever two interpretations of the visual input are mutually exclusive, be it in binocular rivalry, in monocular rivalry, in masking or in bi-stable perception.

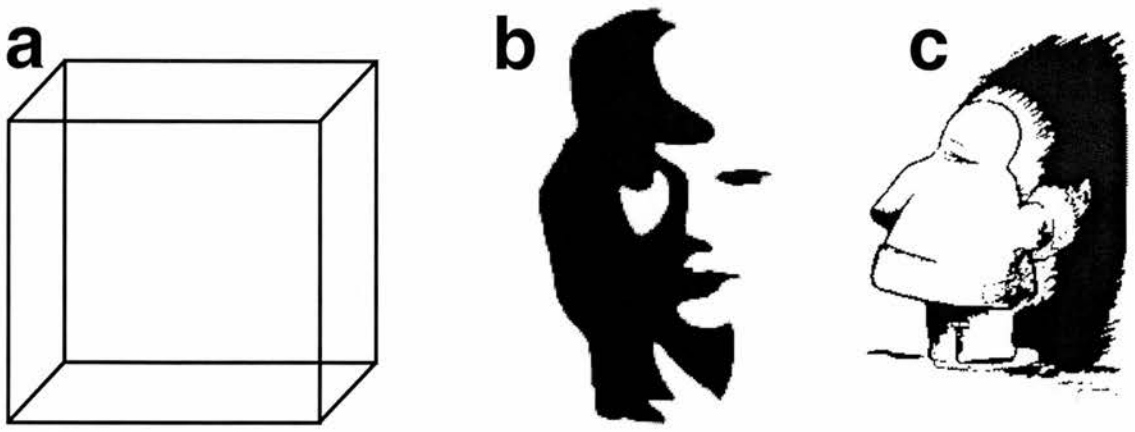


Figure 4.1: (a) Rivalry between two possible 3D interpretations of the Necker cube. If your task is to report which way the cube is pointing, you need to choose between the two alternative perceptions. The decision between the two alternatives appears to occur at the perceptual stage, creating a single perception, and not after perception, with the subject perceiving both interpretations simultaneously. (b) Rivalry between a woman's face and a saxophone player. (c) Rivalry between the profile face of an Indian and the back view of an Eskimo.

Clearly, there are differences between masking, binocular rivalry, monocular rivalry and bi-stable illusions. For instance Kovács et al. (1996) showed that the eye of origin does play a role in binocular rivalry, albeit a surprisingly small role. Also, in masking, some forms of masking could occur before the two ocular channels converge in V1, while such mechanisms cannot be responsible for binocular rivalry. The claim that masking and binocular rivalry are part of a larger class of phenomena that I call 'stimulus competition' does not deny these differences. Instead, it acknowledges specificities, while drawing attention to the overwhelming similarity: in all those cases more than one mutually exclusive stimulus are represented in the visual system, and compete against each other. In both binocular rivalry and masking, competition takes some time: ~60ms.

The aim of claiming that rather than thinking in terms of masking or binocular rivalry, one should think in terms of stimulus competition, is to bring together facts that were previously isolated in the hope that research on binocular rivalry can gain insights from the literature on masking, and the other way around.

In both cases, there is at present little understanding of the neural mechanisms responsible for the competition of the mutually exclusive neural interpretations of the visual input. Until these mechanisms are understood, it will be impossible to know if the same mechanisms are responsible for all cases of stimulus competition. Yet,

considering phenomena together as stimulus competition may accelerate the process of understanding their underlying neural machinery.

The vast literature on attention, which is not going to be reviewed here, has recently been reconceptualised as being a bias in the competition between stimuli (Duncan, 1998; Desimone, 1998): in a complex array of stimuli, drawing attention to one stimulus may favour this stimulus in the competition between stimuli. Compared to masking or binocular rivalry, in a complex array of elements, competition between elements is weaker, and does not result in the perceptual disappearance of one of the elements. This may be, because elements are not mutually exclusive in selective attention paradigms. For masking or binocular rivalry to occur, stimuli (or their interpretation) appear to have to be mutually exclusive. It remains to be understood, how the weaker competition between elements of an array in a selective attention paradigm compares with the stronger competition occurring in masking and binocular rivalry.

In the next Chapter, the RSVP paradigm will be introduced. RSVP too, is to be thought of as temporal competition between stimuli.

5. Introduction to RSVP

In Chapter 3, the literature on visual masking has been presented. Uttal (1969a,b) demonstrated that presenting a mask both before and after a target resulted in detection performance for the target that was inferior to that expected based on the forward and backward masking alone.

Under natural viewing conditions, a saccade occurs approximately 3 times a second (Liversedge and Findlay, 2000). The visual input to be processed by our brain is therefore a continuous sequence of snapshots of the world, falling one after another on the same region of the retinae. Each snapshot is both preceded and followed by other images: a situation in which performance cannot be predicted by measuring performance is simple forward or backward masking paradigms.

RSVP (Rapid Serial Visual Presentation) is a paradigm in which a sequence of individual images are presented one after another on the same location on the screen. The motion pictures that we are all familiar with are a particular case of RSVP. In motion pictures, the still frames of the sequence are successive snapshots of a continuous event, creating the appearance of motion. In RSVP, this is exceptional: most of the time, successive images are not related to each other and they do not create the illusion of motion. RSVP thus resembles more an extremely rapid slide show, in which individual still images are presented one after another in rapid sequence. Movies 5.1 and 5.2 on the attached CD illustrate examples of RSVP sequences. Movie 5.1 is a rapid RSVP sequence, in which each image is presented for approximately 33ms (the NTSC standard is in use) while Movie 5.2 is the same sequence presented at a rate of 264ms/image. As can be seen, the presentation rate of an RSVP sequence determines how perceivable the individual images in the sequence are.

In this Chapter a brief review of the literature on RSVP will be presented. This review will not attempt to be exhaustive but only to create a background knowledge of the basic phenomena associated with RSVP sequences. A more extensive review can be found in the book "Fleeting Memories: Cognition of Brief Visual Stimuli": an excellent collection of sections written by some of the leading experts of the field and edited by Coltheart (1999). This knowledge will set the background for the experimental investigations of the present thesis.

5.1. Memory and detection performance in RSVP

In the mid-sixties, two researchers (Shepard, 1967; Nickerson, 1965), presented hundreds of photographs to subjects at a rate of one image every ~5s. After the sequence, subjects were asked to decide, if a number of images had ('old') or had not ('new') been part of the original sequence. In both studies, performance in this memory task was extremely high: ~90% correct old/new decisions. Standing (1973) presented ten thousand photographs and measured a similar memory performance.

RSVP was developed in 1969 by Potter and Levy (1969) to test how accelerating the presentation process of images would influence the recognition performance. They presented subjects with RSVP sequences of 16 photographs at presentation rates ranging from 125ms/image to 2000ms/image, and tested recognition performance for these images by presenting subjects with 32 photographs mounted on cardboards, 16 of which had been in the sequence and 16 which had not been in the sequence. Subjects were then asked the question "Was this photographs in the sequence?" for each image and had to answer 'yes', 'maybe' or 'no'. The results indicated 3 effects. First, the proportion correct recognitions decreased with decreasing presentation duration per image, being over 90% for 2000ms/image, and 13% at 125ms/image. Second, they showed that the last image in the sequence was recognised much better than all other images (66% percent correct 'yeses' for the last image vs. 13% correct 'yeses' for all others at 125ms/image). This demonstrates that the masking effect of subsequent images is strong. Third, they demonstrated that the probability of correctly recognising item *i* in the sequence is the same whether item *i-1* had been correctly remembered or not.

Later Potter (1976) investigated if the relatively low memory performance for rapid RSVP sequences was due to an inability to perceive the stimuli during the sequence, or to recognise the items after the sequence. She again presented sequences of 16 photographs. In her new study, subjects were split in 3 groups. All 3 groups saw the exact same sequences, but had different tasks to perform. The first group, called 'recognition group' performed the same task as in the previous study: they viewed an RSVP sequence, and afterwards, performed an old/new recognition task on all images in the sequence. This group measured how much recognition memory was available after the sequence. The two other groups served to test how well individual images can be perceived *during* the sequence itself. Trials started by the presentation of a

target, either by projecting the target itself for 500ms ('picture target group') or by projecting a slide that contained a sentence of the type "a street with cars", describing the target image ('name target group'). Figure 5.1 illustrates the performance of the subject in the 3 groups. At a presentation rate of 113ms/image, subjects could detect the target with 64% accuracy (corrected for guessing) in the picture target group, and just slightly less accurately in the name target group. Subjects were thus quite good at perceiving the images during the sequence. The recognition group on the other hand performed poorly, at 11% correct after correction for guessing. It thus appears that subject can perceive images in the sequence, but that they forget what they had seen before their memory can be tested in the old/new recognition task. RSVP appears to prevent memory consolidation.

In a further experiment, Potter (1976) demonstrated that actively looking for a target in the sequence was detrimental to the memory of the other, non-target images in the sequence. She showed, that if subjects had to search for a target, and where then afterwards tested for their old/new recognition memory for the non-target images, their memory was poorer than that of the subjects in the 'recognition group' of the first experiment, which had no search task to perform.

Potter (1976) also showed that RSVP is different from masking. She generated hybrid sequences, which contained the same images as her usual RSVP sequences, but images were separated by a long, 4.5s pattern mask. This hybrid image-mask-image-mask-...-mask sequence differed from a normal RSVP sequence in one fundamental aspect. In RSVP all images are peers: all are equally likely to be tested in the recognition task, all are presented for the same duration, all are complex stimuli. Hence, all stimuli are likely to be processed equally, except if a subject is searching for a particular target, but this is not the case in the standard recognition task. In the hybrid sequence, the mask is not a peer of the target. The masks were clearly different from the other images, and the subjects knew, the mask would not get tested in the recognition task. Hence, subjects could ignore the masks for the requirements of the task. Potter (1976) found, that recognition performance for the images in the hybrid sequences was much higher than in a normal RSVP sequence, with over 40% correct performance (corrected for guessing) for 50ms target images compared to only 11% correct for RSVP sequences at 117ms.

Potter suggests that this difference in performance between normal RSVP and hybrid sequences is central to understanding why in RSVP sequences, items that can

be correctly identified are forgotten so fast. As mentioned, in the hybrid sequences, the mask can be ignored. Potter suggests, that if the mask is ignored, it is not processed at high, 'conceptual' levels of visual processing, and hence, the representation of the previous stimulus at those high-levels is not interrupted by the mask. In a normal RSVP sequence, all stimuli need to be processed equally at the high 'conceptual' levels, and hence, every stimulus interrupts the representation of the previous stimulus at all levels, including the 'conceptual' levels. She finally ventures, that unless the conceptual representation of a stimulus lasts for longer than a certain period, the item will be forgotten. Only pure RSVP sequences will interrupt the conceptual representation of the stimulus before the time necessary for it to be remembered. She calls this effect 'conceptual masking'.

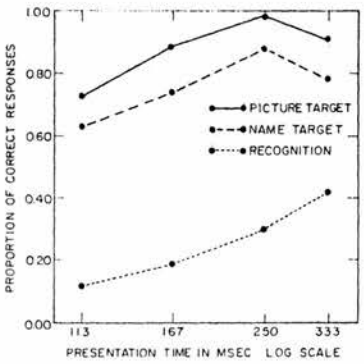


Figure 5.1: Proportion correct responses, corrected for guessing, in the Potter (1976) experiment. The two top curves represent the performance in a search tasks, in which a target is shown, and subjects have to decide if that target is present in a following RSVP sequence containing the target in 50% of the cases. The bottom curve is the performance in an old/new recognition task given to another group of subjects after each RSVP sequence. See text for details. [Adapted from Potter (1976)]

Subramaniam et al. (2000) recently investigated further the inability to remember images in RSVP sequences. They presented RSVP sequences of 32 line drawings at 72ms/image and 126ms/image after a target drawing had been specified verbally. Subjects could easily perform this task, with 77%[5] correct at 72ms/image and 86% correct at 126ms/image. They also measured recognition performance for images at 126ms/image for sequences of 32 line drawings, and measured 58% correct in a two alternative forced choice (2AFC) old/new recognition test. They also tested longer sequences of 92 images, and found the recognition performance to be even worse: 52% correct. Again, this supports the idea that images are correctly identified

[5] All scores in Subramaniam et al. (2000) are not corrected for guessing, and are expected to be 50% correct by chance.

in RSVP sequences at presentation rates ~100ms/image (~80% correct), but that the images are rapidly forgotten, resulting in a chance performance in the recognition task.

Subramaniam et al. (2000) also showed, that at 72ms/image and 126ms/image images do not prime the subsequent detection of the same image at the same presentation rate. They showed this by demonstrating that seeing an image up to 31 times as nontarget in previous sequences did not rise the detection performance of that image when it became a target compared to images that had never been seen as non-targets before. On the other hand, showing the same image a single time outside of an RSVP sequence results in a ~10% increase of detection performance in a subsequent sequence.

How can it be explained that at a given presentation rate, images are correctly identified, but recognition memory of the images is at chance, and no priming of subsequent identifications of the same images occurs? Subramaniam et al. (2000) propose to explain this finding based on two lines of evidence from single cell recordings in the macaque IT cortex.

First, Rolls et al. (1994a,b, 1999) showed that the neural activity following a 16ms target image lasts for several hundred milliseconds unless another image (the mask) is presented shortly after the target. In the latter case, the mask interrupts the neural activity to the target, and restricts the duration of the neural activity to the duration of the stimulus onset asynchrony (SOA, actually SOA+~60ms, see Chapter 3).

Second, Tovee et al. (1993) demonstrate that over 50% of the information about a stimulus contained in the firing pattern of IT cells is contained in the first 50ms of their responses.

Together these findings lead Subramaniam et al. (2000) to extand Potter's hypothesise by stating that only the first 100ms of the response are required for identification of a target, but that single cell responses normally continue for longer, and that this prolonged activity is necessary for the creation of a memory trace. In RSVP, when presentation rate is ~100ms, the response duration of single cells in the brain is limited to ~100ms: a duration sufficient for correct identification, but insufficient for the creation of a memory trace. Recognition performance and priming are dependent on this memory trace and are thus at chance at those presentation rates. The problem with Subramaniam et al.'s interpretation is that Potter (1976)

demonstrated that the lack of memory for stimuli in RSVP sequences is not observed for masked stimuli of the same durations. Nevertheless, Subramaniam et al. explain the lack of recognition memory by using single cell findings collected using a backward masking paradigm. Clearly, a physiological investigation of RSVP would be required as a basis for understanding the neural mechanisms underlying the lack of recognition memory for stimuli in RSVP.

5.2. Attentional blink (AB)

Potter and Levy (1969) demonstrated that if a given item is correctly memorised in an RSVP sequence, this does not influence the memory performance for the directly following item. In their study, subjects had no reason to attend to any particular item in the sequences, and all items thus appeared to be processed equally and independently, except for the last item that was better memorised due to a lack of masking through subsequent images. But what happens, if subjects actively look for a particular image in the sequence?

Raymond et al. (1992) investigated that question. In their first experiment, using the RSVP paradigm, they presented capital letters, one by one, at a rate of 90ms/letter. All letters were black on a grey background, save one, which was white on a grey background. Subjects were instructed to report the white letter, together with the 3 letters that directly followed it. The white target letter was reported 80% of the time, but the letters directly following it were reported no more than 30% of the time each. It appeared as though detecting the white letter prevented the correct identification of the following letters.

In their second experiment, Raymond et al. (1992) had two conditions. In the first, 'experimental' condition, the task was again to name the white letter, but rather than to name the next 3 letters as well, subjects just had to decide, if a capital X, which occurred after the white target letter in 50% of the sequences, had occurred in this particular sequence. In the control condition, everything was identical, except that subjects were told to ignore the white letter, and just to decide if the X occurred. Figure 5.2a illustrates the findings of Raymond et al.: while the white letter created a small masking effect even if it was ignored in the control condition, in the experimental condition, reporting the white target letter resulted in a significant reduction of the probability to perceive the X if the X occurred between 180 and

450ms after the white target letter. It appears, as though identifying the white target letter results in the closing of an attentional shutter that protects the processing of the target letter from other stimuli occuring in the interval 180ms-450ms. Raymond et al. (1992) named this phenomenon 'Attentional Blink' (AB): just like blinking with the eye lid, detecting a searched for target will reduce the probability to see a second target in the interval 180 to 450ms.

In their 3rd Experiment, Raymond et al. (1992) introduced between 1 and 3 blank frames after the white target letter occurred, and tested how these blanks effect the probability of detecting the X in the sequence. Surprisingly, as can be seen in Figure 5.2b, a blank as short as 90ms seems to reduce, and a blank of 180ms seems to abolish the AB effect. This is surprising considering the usual duration of ~450ms for the AB effect in conditions without gaps (up to 450ms). It appears, that like a protective reflex, if a target is followed by a masking stimulus, AB sets in to protect the target from interference from the masking stimulus. AB then lasts for up to 450ms. If on the other hand no stimulus follows the target, AB is unnecessary. After about 100ms, the processing of the target is concluded, and new stimuli can be processed without AB.

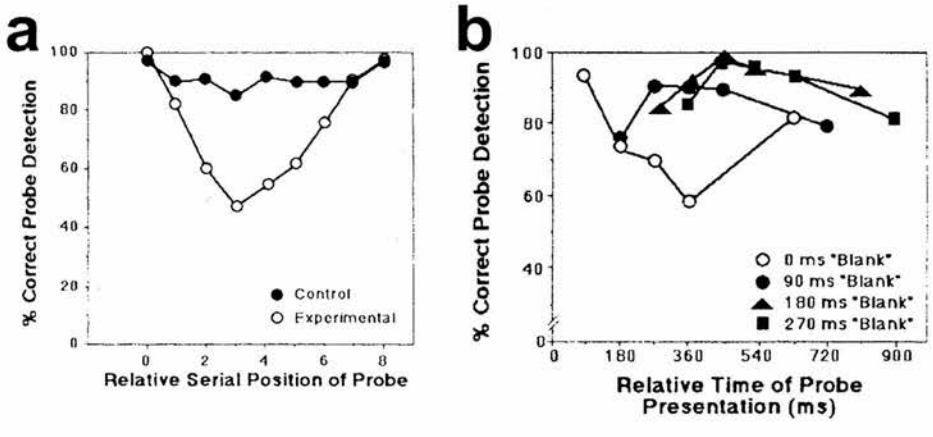


Figure 5.2: The Attentional Blink (AB) effect. (a) Mean percentage of trials in which the X (='probe') was correctly detected as a function of the position of the X relative to the white target letter (position 0). RSVP presentation rate was 90ms/image. The filled symbols correspond to trials in which the subjects were instructed to ignore the white target letter, and the open circles to cases in which the subjects had to attend to and report the white target letter. Clearly, if the X occurred in positions 2 to 5, it was much less likely to be reported if the white target had to be reported, than when it did not. This illustrates how directing attention to the white target letter 'blinked out' the X in the positions 2 to 5 after the white target letter, corresponding to a post-target interval of 180-450ms. (b) Proportion correct X detection as a function of the length of a blank directly following the white target letter. Subject had always to report the

white target. A 90ms gap reduces the AB effect, and a 180ms gap almost abolishes the AB effect. [Adapted from Raymond et al., 1992]

The AB effect is robust, and has been observed by a large number of investigators (Broadbent and Broadbent, 1987; Lawrence, 1971; Chun and Potter, 1995).

5.3. Repetition blindness (RB)

In 1987, Nancy Kanwisher presented subjects with sequences of words, each word being presented as one frame in an RSVP sequence. One word in each sequence appeared twice, and subjects had to indicate which word had appeared twice. The second repetition (R2) of the word was written in lowercase, while the first repetition (R1) was written in uppercase. Subjects also had to rate how confident they were about their response. Presentation rates between 117ms/image and 250ms/image were used. At these rates, subjects are usually quite good at identifying words. Her results indicate that subject in RSVP sequences appeared to be rather blind to the second, repeated occurrence of a word: at 117ms/image for instance, with one word interveaning between R1 and R2, only 43% of the repetitions were correctly reported, and in only 5% of the cases was the correct report made with confidence. Performance increased for both longer SOA between R1 and R2, and longer presentation times per image. In her experiment 2, Kanwisher (1987) presented RSVP sequences of words that formed a sentence at a presentation rate of 117ms/image, and asked subject to repeat the sentence out aloud directly after the presentation. She compared 3 conditions, as shown in Table 5.1, and found that while the first occurrence (R1) of a repeated word is correctly remembered, subjects appeared blind to the second occurrence (R2) of the word and simply omitted R2 from their report, thereby sacrificing the meaning and grammaticality of the reported sentence. If R1 was replaced by a synonym, subject suddenly reported R2, indicating that it is truly the repetition of the word that caused the 'blindness' to it.

Condition	Example sentence	Probability of remembering R1*	Probability of remembering R2**
Repeated	When she spilled the *ink** there was *ink*** all over	.90 [ink]	.22 [ink]
Unrepeated	When she spiked the *liquid** there was *ink*** all over	.94 [liquid]	.79 [ink]
Blank	When she spilled the *ink** there was ** all over.	.92 [ink]	.0 [ink]

Table 5.1: Blindness for a repeated word. For all three conditions, an example of a sentence is shown, followed by the probability of recalling the correct (as indicated in the square bracket) word for the first position of the sentence, masked by a * in the example sentence, and the probability of recalling the correct word for the second position (**). Clearly, the verbatim repetition of a word is very frequently omitted. Note that the * and ** were not presented with the sentence. Adapted from Kanwisher (1987)

Why do people omit the second repetition of a word? Kanwisher believes that full perception of a word is composed of two components: the 'type' and the 'token'. If you have to decide how many pears are in a fruit bowl, you have to differentiate pears from apples, which are different 'types'. Then you also have not to consider all pears as one, but you have to know that there are for instance 5 examplars of the type pear. This, in Kanwisher's model, is done by giving each exemplar a 'token'. If in RSVP subjects fail to report R2, it is not because they did not perceive R2 but because they fail to differentiate R2 from R1. In Kanwisher's terminology: subjects failed to give R2 a token separate from R1 which shared the same type.

Could it be that detecting R1 fatigued the neurones that are responsible for its detection, and that R2 fell into a refractory period of the neurones, which prevent its processing? Kanwisher (1987) showed that is was not the case. She designed a slightly different task, in which the list of words ended with R2, followed by a mask. Similarly to the hybrid sequences of Raymond et al. (1992), R2 was no longer masked by a peer stimulus, and thus became itself more prominent. Under these conditions, R2 was detected *more* often if the same word occurred earlier in the sequence – just the opposite of what would be expected if R1 caused a refractory period. This effect commonly known as repetition priming, demonstrates that the system is quite capable of recognising R2: the problem is to give it its 'token'. Presenting it at the end of the sequence makes it more dissimilar from R1, and helps the process of giving R2 a separate token.

From a neurophysiological perspective, Kanwisher's idea of token can be understood in the context of neural persistence. Brief stimuli produce neural responses

that outlast the duration of the stimuli. If the stimulus is presented again briefly after the first occurrence, the activity caused by R2 will create a problem for the brain. Is the activity caused by R2 evidence for a new stimulus, or is it just the persisting response to the first stimulus? Deciding that it is a new stimulus is comparable in Kanwisher's model to give it a new token, deciding that it is persisting activity is equivalent to displaying RB and to fail to give it a new token.

The effect of RB has been replicated in a number of studies, and has been demonstrated for a variety of stimuli: words, letters (Kanwisher and Potter, 1990), colours (Kanwisher 1991) and object pictures (Bavelier, 1994; Kanwisher and Yin, 1993). RB depends on attention: if R1 is ignored, R2 will not be missed (Kanwisher 1991).

5.4. RSVP and saccadic exploration

RSVP has received much interest in the psychological literature. This is partly because of its value as a model for saccadic exploration of a natural scene: in both cases the visual input is a continuous sequence of often relatively still images. Despite this similarity, a number of differences between RSVP and saccadic exploration are worth discussing.

First, for saccades, the object that is going to be foveated after the saccade is not unpredictable: it was usually present in the peripheral visual field before the saccade. Indeed Moore et al. (1998) demonstrated, that during a visual saccade, V4 cells with receptive fields in the periphery show a burst of activity to a stimulus that had been in their receptive field before the saccade, if this stimulus is the target of the saccade, but much less, if another location is the target of the saccade. Hence, before the 'new'[6] stimulus reaches the fovea, its peripheral representation will be strongly active, and may facilitate or prime the processing of the same stimulus at its new, foveal input location. In RSVP, such priming is generally impossible because the preceding image did not contain the same stimulus in the periphery. Even if the preceding image contained the same stimulus as the new image, but at a different location, the mechanism demonstrated by Moore et al. depends on the saccade being

[6] New is placed in inverted commas because the stimulus may be new to the fovea, but had been previously part of the peripheral visual field.

aimed at the location at which the stimulus was in the preceding image, and thus would not operate in the context of RSVP.

Second, a saccade is the result of a motor command initiated by the brain. This motor command is available to the brain in natural saccadic exploration, and is time locked to the change of the visual input. It could be used by the brain to help process the sequence of visual input. It is known, that during a saccade, visible events are hard to detect, a phenomenon called 'saccadic suppression' (see Volkmann, 1986 for a review). The 'corollary discharge' hypothesis states that an efference copy of the oculomotor command is used to inhibit visual neurones during a saccade, causing the 'saccadic suppression'. An alternative flavour of the same basic idea states, that not the motor command, but the proprioceptive ocular signals resulting from the saccade are the source of the inhibition. If either of the flavours was true, RSVP would be quite different from saccadic exploration in that in saccadic exploration, the visual input would be parsed in separate entities by the inhibitory corollary discharge associated with the oculomotor command. In RSVP this parsing signal would be absent, and stimuli might interfere more strongly with each other if they are not separated by this inhibitory discharge.

Wurtz (1969a,b) tested the corollary discharge hypothesis by recording in V1 while (a) a monkey saccades, and a stimulus is placed in the trajectory of the saccade so as to sweep through the receptive field of a V1 cell ('eye motion' condition) or (b) the monkey's eyes were stationary and the same stimulus was swept through the receptive field at a velocity comparable to that caused by the eye movements in (a) ('stimulus motion'). The corollary discharge hypothesis would predict, that single cell responses should be quite different in those two conditions: the corrolary discharge should cause inhibition during eye motion but not stimulus motion. Wurtz (1969a,b) showed that this was not the case: single cells in V1 responded equally to eye motion and stimulus motion. Yet, only 32% of all cells responded to stimuli moving at saccade velocity (~900deg/s), the other 68% which respond with an excitatory response to a slowly moving bar, stopped responding to stimuli moving at saccadic velocity. Hence, the result suggest that saccadic suppression is more likely to occur due to the fact that single cells in V1 do not readily respond to stimuli moving at saccadic velocity.

Later studies by Judge et al. (1980) demonstrate that V1 neurones which respond during eye motion, do not respond, if an effective stimulus is turned on in

their receptive field before the saccade. Presenting an effective stimulus in their receptive field after the saccade had a much smaller effect. These findings indicate that in V1 cells, the responses occuring during a saccade can be masked by visual input before the saccade, and to a lesser extend, after the saccade. Together with the earlier findings of Wurtz (1969a,b) showing that responses are equal during eye motion and stimulus motion, this indicate that 'corollary discharges' do not account for saccadic suppression of V1 responses. Instead, purely retinal sources of information appear to be responsible for saccadic suppression, namely (a) stimulus velocities at which few neurones respond, and (b) forward masking through the static image before the saccade. The forward masking may further be increased by the reactivation of responses to the stimulus towards which a saccade is directed (Moore et al., 1998).

The fact that there is no evidence for eye movements having a direct impact on the response of V1 cells beyond those explained by the visual input itself, supports the idea that the neural processing involved in saccadic exploration is not that much different from the processing in RSVP. The only difference would be the existence of a blurry, fast moving visual input during the saccades in saccadic exploration, while this retinal input is absent in RSVP. Yet during a saccade, a minority (1/3) of neurones are activated by visual stimuli, and this signal only occurs for about ~30ms every 300ms. Actually, RSVP sequences could be constructed to contain such a ~30ms blurry episode between two consecutive images, but this will not be done in the present thesis. RSVP in this thesis will be used as a model of visual processing freed from the limitations of the body: as if saccade velocity was infinite.

Results obtained using the RSVP paradigm are thus useful indication of what may happen during active saccadic exploration, especially since RSVP-type masking appears to be critical for saccadic suppression: an element responsible for the stable perception we have of the world. RSVP is different from and better than masking (see below) when it comes to understanding saccadic exploration. Yet, there are differences between saccadic exploration and RSVP: in RSVP there is no priming through the peripheral visual field, and there are no ~30ms blurry visual inputs separating consecutive images. Thus, while RSVP can help us understand visual processing during saccadic exploration, the aforementioned differences make RSVP more challenging for the visual system than saccadic exploration, and should be kept in mind.

5.5. Is RSVP different from Masking?

One of the aims of the present thesis is to investigate the response of single cells in the STSa during RSVP. Rolls et al. (1994a,b, 1999) and Kovács (1995) measured the responses of single neurones during backward masking in IT and STS. This begs the question: is it necessary to replicate these investigations for RSVP? Is RSVP not likely to be identical to masking at the cellular level?

Given that RSVP has so far not been investigated at the single cell level, it cannot be decided if RSVP will be identical to backward masking at the cellular level. There are nevertheless many reasons to believe that it will not: masking and RSVP differ in their impact on perception and memory, as measured at the behavioural level.

As mentioned earlier, Uttal (1969a,b) demonstrated that placing a mask both before and after a target makes the target more difficult to detect, than expected based on the summed effect of forward and backward masking alone.

Potter (1976) demonstrated that memory for masked items is much better than memory for the same items presented for the same duration in an RSVP sequence. This difference is thought to arise from the fact that in RSVP, all items have an equal status, and all have to be processed. The processing of a new stimulus is thought to interrupt the representation of the preceeding stimulus at all levels of processing. In masking on the other hand, the mask is clearly different from the target. While the mask probably interrupts the low-level, visual representation of the stimulus, higher level, more semantic processes have no reason to process the mask, and thus can continue to process the target in a masking paradigm even after the mask has been presented. It is thought that the continued activity of those higher levels create the stronger memory trace in masking compared to RSVP. The findings of Uttal demonstrate that in addition, the detection of a stimulus is detrimentally affected by the combination of forward and backward masking, which will add to the superiority of performance in backward masking vs. RSVP.

Similarly, Subramaniam et al. (2000) demonstrate that no visual priming occurs in RSVP sequences at presentation rates of 126ms/image. Backward masked stimuli do produce visual priming at the same SOA. This finding confirms the difference between RSVP and masking.

Finally, Kanwisher (1987) demonstrates that repetition blindness occurs in RSVP sequences but not in masking. If a word is embedded in an RSVP sequence, i.e. is followed by other words, then that word will be remembered less often, if earlier in the sequence the same word had occurred, than if it had not. On the other hand, the target word occurs last in an RSVP sequence, and is just followed by a pattern mask that is not another word, the word is remembered *more* often, if the same word occurred earlier in the sequence. Hence masking a stimulus is different from embedding it in an RSVP sequence

Taken together, these studies clearly indicate differences between RSVP and backward masking at the same SOA, and are further reasons justifying the need for a physiological investigation of RSVP.

5.5. Summary and conclusions

In RSVP a number of images are presented one after another, at a rapid presentation rate. Increasing the presentation rate has detrimental effects on both perception and memory. If subjects are asked to search in an RSVP sequence for a particular target, their performance will decrease with increasing presentation rate. If subjects are asked to remember items in the RSVP sequence for a subsequent old/new recognition test, their performance will also decrease with presentation rate. Memory for items in RSVP sequences is much poorer than memory for the same items under backward masking conditions, suggesting that the other meaningful images in an RSVP sequence prevent the consolidation of a memory trace for items, which can be correctly identified. Recognition memory for items in an RSVP sequence is poorer, if a subject was actively looking for another item in the sequence, than if the subject had no particular task.

Detecting a target in an RSVP sequence is detrimental to the detection of subsequent images occurring ~200-500ms after the target, but only if the first target is difficult to detect, and the subject was actively searching for that target, a phenomenon called attentional blink.

If an item is repeated in an RSVP sequence, the second occurrence of the items appears to be misattributed as visible persistence of the first occurrence of the item. As a consequence of this attribution, subjects appear to be blind to the second occurrence: a phenomenon called repetition blindness. Repetition blindness increases

if presentation rate is increased and if the two repetitions are closer in time to one another. Repetition blindness extends for several hundred milliseconds after the first occurrence of the word.

RSVP is different from masking, and the large number of psychological investigations using RSVP calls for a physiological investigations of single cell responses during RSVP.

The next Chapter will start the experimental section of this thesis by presenting such a physiological investigation of single cell responses during RSVP presentation.

6. General Electrophysiological Methods

Because RSVP has not been used in the past to investigate single cell response in the higher visual cortex, the methods are new, and, therefore, may benefit from being described in a dedicated chapter.

6.1. Subjects and surgery

Two adult *Macaca mullata* ('Steve' and 'Terry') with a weight of 12.4kg and 8.5kg were used. Two recording chambers were implanted previously in both monkeys under fully aseptic conditions and general anaesthesia. In addition, two plastic tubes were placed in front and behind the recording chambers, perpendicular to the saggital plane to restrain head movements. The stainless steal recording wells and the plastic tubes were held in place by dental acrylate anchored to the scull by stainless steel screws and "T" pieces. After implantation, the skin was closed around the dental acrylate implant, and lids were used to close the recording wells. Every two to 4 days, the recording wells were opened, rinsed with saline, and the lids cleaned with disinfecting soap. All procedures were performed according to U.K. Home office protocols, and under a Home office project license. All experimenters held a home office personal license for all procedures performed. Monkeys were kept under light water deprivation, but had ad lib access to dry food supplemented by fruits and vegetables.

6.2. Recording procedure

The general recording procedures were standard. The monkey was seated in a primate chair and his head was restrained using two metal rods passed through the plastic tube in the monkeys head implant. Lignocaine hydrocloride (Xylocaine 40mg/ml) was applied topically to the dura. Using a David Kopf micropositioner fixed to the recording well, a guide tube was inserted transduraly to a depth of approximately 2mm relative to the dura. A glass coated tungsten electrode was then passed through

the guide tube and lowered manually first, then using a hydraulic microdrive onto the anterior section of the Superior Temporal Sulcus (STSa). Recording positions were based upon locations in which in former experiments face selective neurones had been found. The electrode was lowered under acoustic control until single neurones were isolated. The target area was recognised by the visual response of the cells and the typical stratification of quiet zones preceding the STS in the path of the electrode. The target area is typically reached 28-32mm after the dura. This long path gives a firm support to the electrode allowing the relatively long lasting stable single cell recordings required for this study. At the end of each recording track, a frontal and a lateral X-ray was taken to allow reconstruction of the recording site relative to landmarks on the scull.

One of the two subjects is still being used for further experiment. For the other monkey (Steve), on the last day of recording, an electrode coated with DiI (1, 1'-dioctadecyl-3, 3, 3', 3'-tetramethyllindocarbocyanine perchlorate, Molecular Probes Europe BV) was used to record single cells from a typical recording location. This fluorescent tracer has been used previously to mark electrode tracks (Snodderly and Gur, 1995; Keysers et al., 2000). Single cell recordings using usual methods were performed using this marked electrode until a face responsive neurone was identified at a recording depth associated with the upper bank of STSa based on the stratification of quite and active zones. A micro-lesion (40 microamps for 40s) was made at that position followed by a frontal and a lateral x-ray. The electrode was left *in situ,* while the monkey was sacrificed, perfused and decapitated. An MRI was taken of the skull with the electrode *in situ*. The electrode was removed from the brain, the brain was removed from the skull, and cryoprotected in sucrose solution. 25μm frozen sections were cut, and two consecutive cuts were kept every 250μm, and mounted on slides. One was immediately inspected under a fluorescent microscope to localise the DiI marking, while the other slice was Nissl stained, to identify the position of the electrolytic lesion. Annex 1 illustrates the convergent result of the DiI tracing and the electrolytic lesion localisation. The position was confirmed using the MRI. Based on the x-rays taken at the end of each recording day, the position of all other electrode tracks in Steve were known to be in the close vicinity of this last electrode recording track. Consequently, all other recording positions were in the close vicinity of the electrolytic lesion apparent in Annex 1, and could be attributed to the upper bank, the

lower bank and the fundus of the STS ranging from anterior coordinates +18mm to +14mm approximately relative to the interaural plane.

6.3. Behavioural training

To ensure that the presented stimuli are actually foveated by the animal, the monkey was trained to fixate a central fixation point on the screen in front of him. This procedure was facilitated by the prior training of the monkey for a different experiment, in which he was rewarded for licking if a green LED was presented in front of him, but discouraged to lick if a red LED was presented, with fruit juice and mildly salty solution respectively.

For the specific requirement of our experiment, the monkey was subsequently trained to fixate a yellow fixation dot on a 20 inch computer monitor placed 50cm in front of him. After a warning tone, the monkey's gaze had to foveate (±5° of angle) a yellow dot within 2s for liquid reward to be delivered to the mouth. Reward delivery was continued as long as the monkey's gaze remained in the window. After a maximum of 10s or whenever the monkey's eyes left the fixation window for more then 200ms, the yellow dot vanished, reward delivery was discontinued, a pause of a few seconds (typically 5s) was introduced and the next trial began. After each session lasting no more than 6 hours, the monkey was returned to the home cage and given access to a bottle of water containing the difference between the liquid obtained in the training/experiment and his water requirement (1*l*). Water was then taken away over night until the next training session.

6.4. Measurement of eye movements

The eye movements were noninvasively measured using an Iview eye tracking system composed of a CCD infrared camera filming one of the monkey's eyes (see Fig. 6.1), a dedicated IBM PC compatible computer and the main computer controlling the experiment. The eye-tracking computer extracted uncalibrated x-y coordinates of the pupil centre by calculating the centre of mass of the dark area defining the pupil in contrast to the brighter iris. The co-ordinates were then outputted as a voltage for x and a voltage for y co-ordinates at a 50Hz sampling rate (dictated by the CCD camera).

the main computer controlling the experiment. The eye-tracking computer extracted uncallibrated x-y coordinates of the pupil centre by calculating the centre of mass of the dark area defining the pupil in contrast to the brighter iris. The co-ordinates were then outputted as a voltage for x and a voltage for y co-ordinates at a 50Hz sampling rate (dictated by the CCD camera).

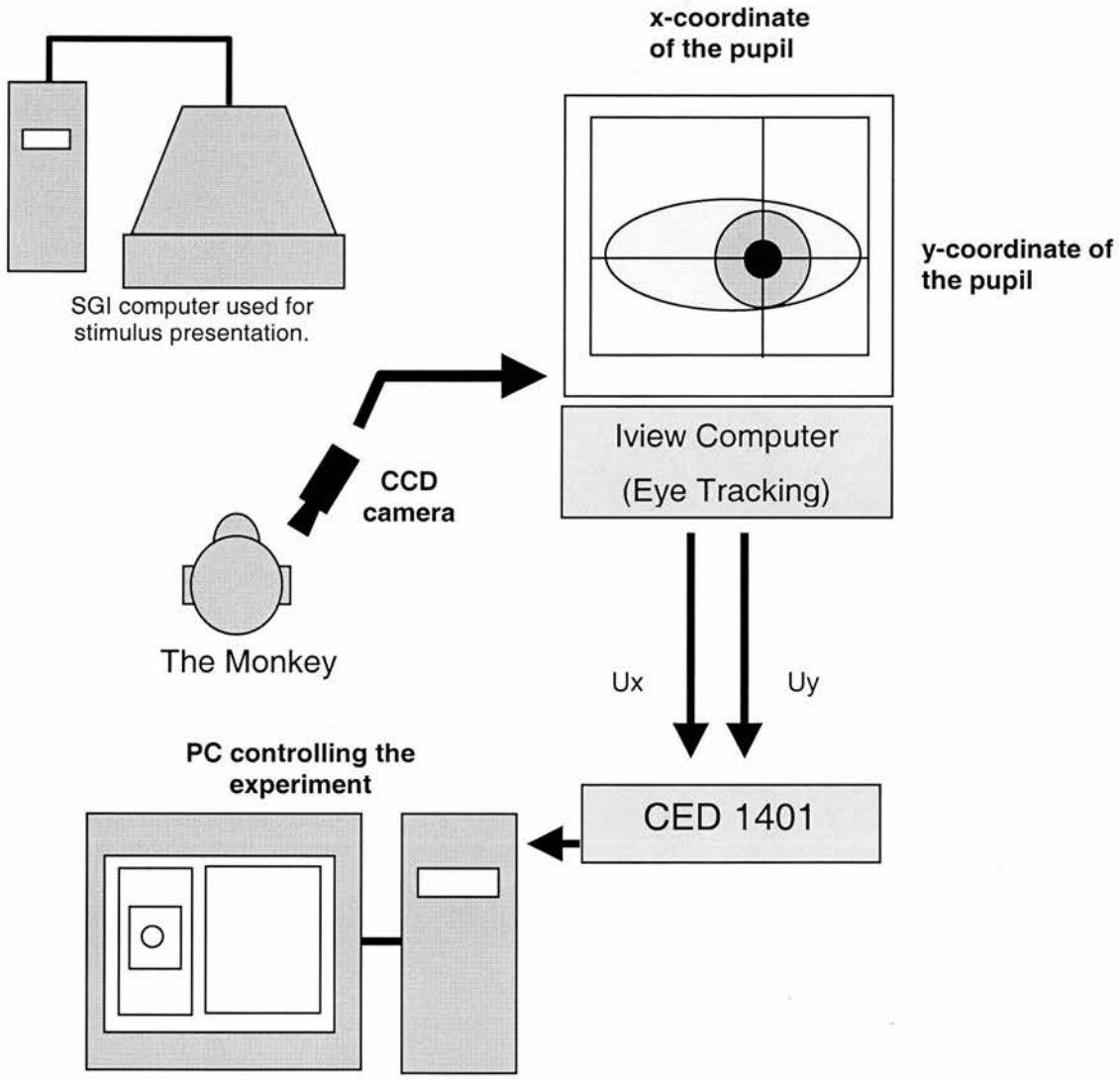


Figure 6.1: Diagram of the eye tracking apparatus. An infrared CCD camera filmed the eye of the monkey while the monkey is looking at the monitor in front of him. The Iview eye-tracking computer calculates the position of the centre of the pupil, and sends an analogue voltage (U_x for the x position of the pupil and U_y for the y position) to the PC controlling the experiment, which converts the voltage through a CED 1401 into eye position relative to the monitor of the silicon graphics workstation (SGI) on which the stimuli are displayed

6.4.1. Calibration

A 5 point calibration was used to convert the X/Y coordinates of the pupil into gaze direction. The problem was to get the monkey to fixate 5 points on the screen (centre + 4 corners). We first unsuccessfully tried to present dots, and later photographs in these positions to attract the monkey's gaze. We finally resorted to a new, very successful technique. Using the video-capture board of the Indigo-2 SGI (Silicon Graphics, Inc.), we streamed the Walt-Disney cartoon "Bamby" in very small format (2° x 2°) to the 5 positions on the screen. The moving pictures of the cartoon seemed to have on the monkeys the same effect it has on children: irresistibly attracting their attention. The monkeys found the movies intrinsically enjoyable to watch, and engaged in sustained fixation without any training. We presented about 10 seconds of the movie in each position, averaged the eye position during that time, and repeated those points that had resulted in too much variation of eye positions during the fixation. This method remains effective even after months of recordings, making a full eye calibration at the beginning of each recording session a quick and effortless procedure. The final precision of the eye tracking system was around 0.5 to 1° of angle. This limited precision is in part due to the size of the movies, but is sufficient given the size of receptive fields in STSa.

6.5. Stimulus presentation

As illustrated in Fig. 6.2, the stimuli used in this experiment are digital photos and pictures from various sources that have been converted to a size of 320 x 256 pixels (extending 10° horizontally and 8° vertically) and are presented centrally by an Ingido2 Silicon Graphics Incorporated (SGI) workstation on a Sony GDM-20D11 monitor with a monitor refresh rate of 72Hz (=13.89ms/Frame). The PC controlling the experiment informed the SGI machine what images to present and for how long each image has to be presented. The PC checked the presentation time through a light sensitive diode that measures the occurrence of a white square on the middle of the sides of the screen. For presentation durations of more than 2 frames/image, this white square is flashed together with the first and last frame of each stimulus, enabling the measurement of start and end time of each stimulus. For 2 frame stimuli, the white square was flashed with the first frame of each stimulus only. For 1 frame

presentations, the white square was flashed with every other stimulus. The PC controlling the experiments stored the input from the diode along with the time of neural spike events in physiological experiments, and discarded all trials where the measured stimulus duration differed from the intended stimulus duration, which happened very rarely (less than once every 50 fixations).

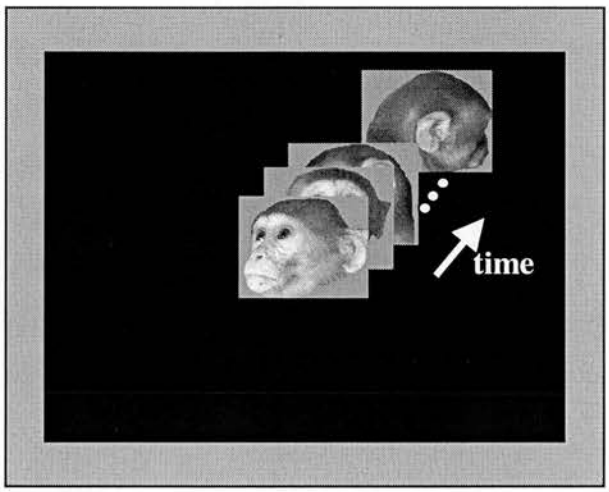


Figure 6.2: Stimulus presentation on the computer screen. The images, coming one after another, are represented offset, one after an other in this illustration, but are all replacing each other in the same location on the screen in the experiment. The "..." represents the many other stimuli being presented during a single fixation.

6.5.1. Duration of an image on a computer screen

As pointed out by Bridgeman (1998), on a computer screen, presentation time of an image is not equal to $1/f$ (where f is the refresh rate in Hz and in 72Hz in our experiment). The resolution of the screen is 1280x1024, where 1024 is the number of lines in each frame. Each frame is drawn line by line, with each line taking 1/1024 of the frame duration to be drawn. Each full frame taking $1/f$ to be presented, each line thus takes 1/72 x 1/1024=14μs to be drawn on a 72Hz monitor. Since our stimuli were only 256 lines high, they are drawn in 3.47ms. The phosphor decay time to 10% intensity for the monitor we used depends on the colour, and is of 1ms for red, 40μs for green and 30μs for blue (Sony, personal communication). Hence, from beginning of the first line drawn to the end of the red decay of the last line of each stimulus, 4.47ms elapse. After this, the monitor remains black for 1/72s-4.47ms=9.4ms. If an image is presented for n frames, its duration is hence not n/f. The true duration, measured from first line of the first frame to the decay of the last line of the last frame

is $(n\text{-}1)/f + 4.47$ms. The ISI between two consecutive frames is 9.4ms, and even less if the image had a zero red content. As seen in our review of the pattern masking literature, the SOA is often a critical value. SOA is simply given by n/f (where f=frame rate in frames/s), since their will be exactly 1/72s=13.9ms between the time the first line of one stimulus is drawn and the first line of the next stimulus at that location on the next frame. If k frames of gaps (=blank screen) are introduced between two images, this will increase the ISI by k/f. The resulting stimulus parameters as used in our experiments are indicated in Table 6.1:

Stimulus [Frames]	Gap [Frames]	Stimulus duration [ms]	ISI [ms]	SOA [ms]
1		4.5	9.4	13.9
2		18.4	9.4	27.8
3		32.3	9.4	41.7
4		46.1	9.4	55.6
8		101.7	9.4	111.1
16		212.8	9.4	222.2
1	3	4.5	51.1	55.6
2	6	18.4	92.8	111.1

Table 6.1: Stimulus duration, ISI and SOA for some RSVP conditions used in Experiments 1-4 as a function of stimulus frame numbers and gap frame numbers. Gap frame numbers refers to frames of black screen presented between two images. Conditions without gap will usually be referred to by their SOA, rounded to the next ms.

6.6. Experimental procedure

For each isolated neurone, we first run the "RSVP search" procedure, and if suitable stimuli were found, we run the "RSVP test" procedure, which was finished either when we lost the cell, or when we had a minimum of 24 trials in the 111ms/image condition.

6.6.1. RSVP search procedure

Stimuli: We used two collection of 30 images each (see Annex 2).

Presentation rate: 111ms/image, no gap

Number of trials: typically 5-10 trials, until we were satisfied to have a reliable estimate of responses

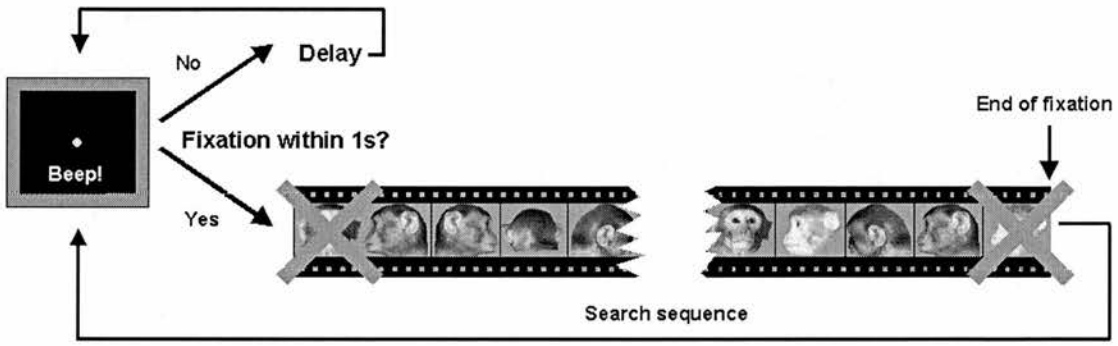


Figure 6.3: Sequence of events in a single trial of the RSVP procedure. A fixation dot is presented in the middle of the screen together with a warning tone ("beep"). The monkey has 1 second to engage fixation within ±5° of the fixation dot. If he fails to engage fixation, the trial is aborted, and after a delay (typically 5 seconds), a new trial is initiated. If fixation is engaged, the stimuli are presented in the centre of the screen, replacing each other in the same location on the screen, until the monkey brakes fixation for more than 200ms (continuous sequences are represented by the celluloid strip. The white rip in the middle indicates the many stimuli not shown). The next trial is then initiated within 2s. The single cell data recorded during the sequence is analysed by the PC and the date from the first and last image during each fixation are discarded because of their privileged position in the sequence. During the entire period of fixation the monkey is rewarded with fluid delivery (fruit-juice or exceptionally coca-cola).

Each trial begins with a tone and a yellow fixation dot (see Fig. 6.3) on the screen in front of the computer. If the monkey engaged fixation within 1s, the fixation point disappeared, and we started image presentation. Images were presented as a continuous flow of stimuli. Each image was presented for 8 frames, and the 9th frame was the first frame of the next stimulus and so on. An image duration of 8 frames (111ms) was used, because this relatively long presentation rate yielded responses that had the same characteristics as those obtained using traditional, single stimulus presentations. The order of the images in the sequence was random with the following restrictions. The continuous flow was composed of blocks of 30 images. In each block, all 30 images were presented in randomised order. The next block had again all

30 stimuli, but in a different order. This ensured, that a stimulus is almost never followed by itself, except in seldom cases, were a stimulus appears last in one block, and first in the next. It also ensures that if recording is interrupted at any point, the same (±1) trial number has been collected for each stimulus. Image presentation stopped after 10s or when the monkey's gaze left the ±5° fixation window for over 200ms – whatever came first. In the later case, data regarding the last stimulus presented before gaze exited the window is discarded. The procedure was continued until enough trials were collected to have a reliable assessment of responses to the different stimuli. Responses were visualised on-line as peri-stimulus time histograms (PSTH) and rastergrams on the computer screen. The experimenter continuously inspected the PSTH. If at any point it became apparent that the cell responded to none of the stimuli, the data acquisition was aborted, and a different search set was used to repeat the procedure. If still no effective stimuli were found, a new cell was isolated. If on the other hand, one of the stimuli was associated with a clear response in the PSTH, data acquisition was stopped to initiate the RSVP test procedure. To this purpose, a test set of 8 stimuli was constructed to span the response range of the cell. All 30 stimuli were rank ordered based on their spike count in the period 100 to 211ms after stimulus onset. This window was arbitrary, and based on an estimated latency of 100ms for STS cells, and a response duration estimated by the stimulus duration (111ms). Since at that time, the latency of the cell was unknown, and we had no knowledge of response duration in RSVP, this window was our best estimate. Of the spike count rank ordered stimuli, the 2 best, 2 worst and 4 intermediate stimuli were selected for the test set. The 4 intermediate stimuli were chosen without any structured rule: if for instance the 2 best and 2 worst stimuli were faces, other face stimuli were usually taken as intermediate stimuli, otherwise, the intermediate stimuli were chosen pseudorandomly to sample the whole range of effectiveness.

6.6.2. RSVP test procedure

Stimuli: A custom tailored set of 8 stimuli was used to span the response range of each individual cell. The stimuli were selected based on the results from the RSVP search procedure: the 2 best, two worst and 4 intermediate stimuli for each cell.

Presentation rate: 6 conditions without gaps: 222ms/image-14ms/image and 2 conditions with gap: s28g83 (i.e. 28ms stimulus, 83ms gap) and s14g42.

Number of trials: We equated the total presentation time in each condition, so that 4 times more trials were collected at 28ms/image than at 111ms/image. On average 350 trials were tested per stimulus in the 14ms/image condition.

Using the test set of 8 stimuli constructed in section 6.6.1., the response of the cells that had been subjectively judged by the experimenters to shown a response to at least one of the stimuli in the RSVP search procedure was assessed at different presentation rates. All cells were tested with a "core" set of presentation rates (see Table 6.2). As the first cells were tested, it became apparent, that a number of additional conditions would be interesting, and these were added to the protocol, and tested in a subset of cells (see Table 6.3).

The different presentation rates were interleaved in blocks (see Table 6.3), so that n_1 trials are gathered for each stimulus at the first rate, then n_2 at the second rate etc., then the whole procedure with blocks of each test rate is repeated again a number of times, until the cell was lost or enough trials were collected. The number of trials collected at each presentation rate before switching to the next are chosen so that the same amount of presentation time is recorded in each condition. (i.e. at 28ms/image 4 times as many trials were run than at 111ms/image). The presentation rates were randomly ordered once, and the same random order was used each time. The conditions of the core experiment were always interleaved with each other. For technical reasons, in 10 cells, the data for the supplementary conditions were interleaved amongst each other, but collected separately, after the core experiment was concluded. In those 10 cells, first the core experiment was performed a satisfactory number of times, and then, after assuring that the cell's response was unchanged, the 3 supplementary conditions were collected, interleaved with each other, but not interleaved with the core experiment.

6.6.3. How many cells were tested with how many trials in how many conditions?

We isolated 137 cells. For 103/137 cells, either no effective stimulus could be found (~70% of cases) or the cell could not be recorded for long enough to complete testing (~30% of cases). In 34 of them, we found selective responses to one of our stimuli and had sufficient time to finish the core experiment (see Table 6.2). In approximately the last 20 of these 34 cells, we also tested the supplementary conditions (see Table 6.3).

Label	Stim [Frames]	Gap [Frames]	Stimulus duration [ms]	ISI [ms]	SOA [ms]	Trials per block per stimulus	Cells tested
222ms/image	16	0	212.8	9.4	222.2	3	34
111ms/image	8	0	101.7	9.4	111.1	6	34
56ms/image	4	0	46.1	9.4	55.6	12	34
28ms/image	2	0	18.4	9.4	27.8	24	34
s28g83ms	2	6	18.4	92.8	111.1	6	34

Table 6.2: Conditions in the core experiment performed on all 34 cells tested. Label refers to the way the conditions will be refered to throughout the paper. Trials per block per stimulus describes how many trials are run for each stimulus before switching to the next testing speed. Once all conditions are tested, the whole procedure is repeated.

Label	Stim [Frames]	Gap [Frames]	Stimulus duration [ms]	ISI [ms]	SOA [ms]	Trials per block per stimulus	Cells tested
14ms/image	1		4.5	9.4	13.9	48	22
42ms/image	3		32.3	9.4	41.7	24	23
s14g42ms	1	3	4.5	51.1	55.6	16	21

Table 6.3: Supplementary conditions tested on only a proportion of the cells as indicated.

Condition	# cells tested	Numbers of trials per stimulus		
		Mean	Minimum	Maximum
222ms/image	34	24	13	42
111ms/image	34	48	27	84
56ms/image	34	95	60	168
28ms/image	34	190	120	315
s28g83ms	34	47	26	84
14ms/image	22	359	48	528
42ms/image	21	126	42	264
s14g42ms	23	138	96	224

Table 6.4: For each condition, the number of cells tested, as well as the mean, minimum and maximum numbers of trials tested per cell.

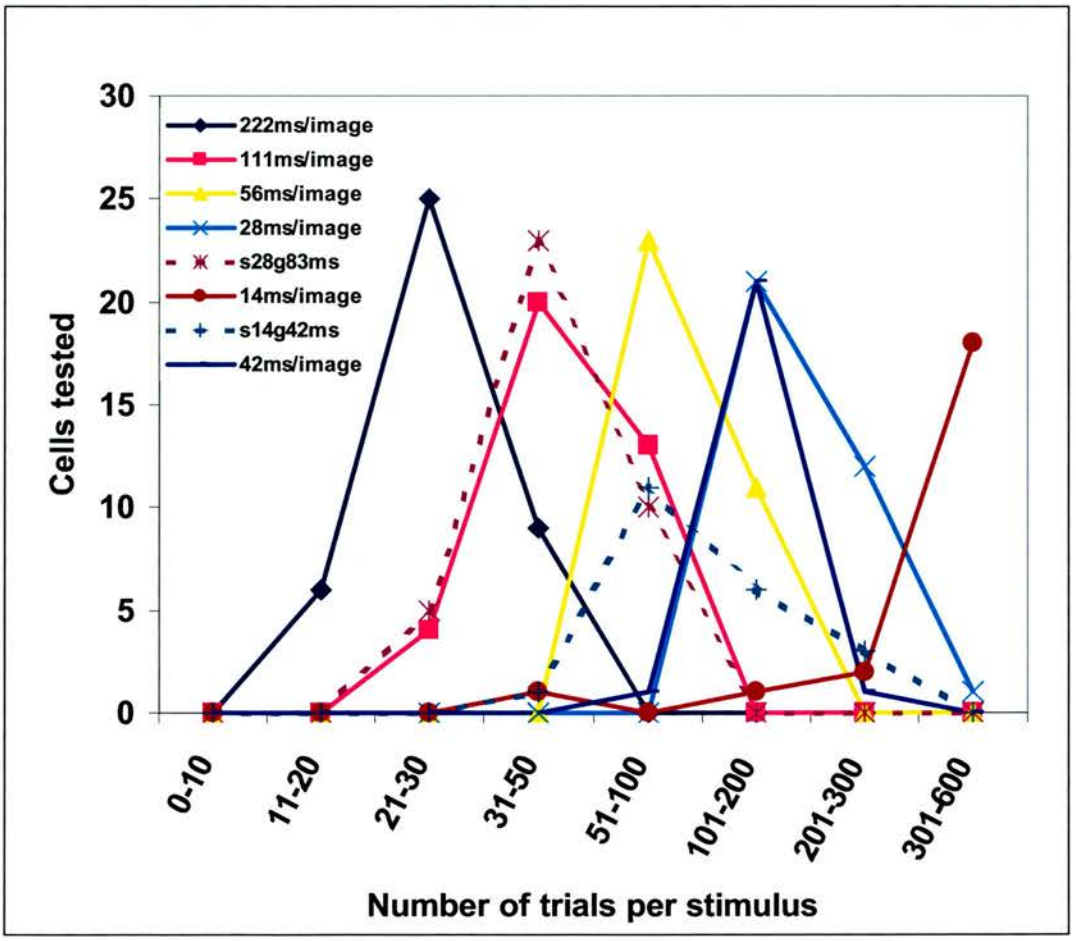


Figure 6.4: Number cells tested as a function of condition and trials tested. Trials with gaps are represented by dotted lines. As can be seen, always at least 10 trials were performed on each stimulus, and the vast majority of conditions contain more than 20 trials.

The RSVP procedure allows the efficient and rapid collection of trials. As can be seen in Table 6.4 and Fig. 6.4, a relatively large number of trials were acquired for each stimulus in each condition. This will be particularly important for information theoretical analysis of the data, and will contribute generally to the statistical power of the data.

6.7. Analysis of the electrophysiological data:

6.7.1. Peri-stimulus time histograms (PSTH)

In RSVP physiology, a continuous stream of nerve responses are collected while a continuous stream of images is presented. Which response is caused by which stimulus? To measure the response to a particular stimulus in the sequence, we

created peristimulus rastergrams by realigning the continuous recording on the time of each occurrence of the stimulus in the sequence. The rastergrams are then used to compute peristimulus time histograms, which are smoothed by gaussian convolution to obtain a spike density function.

In traditional paradigms, after a period without visual stimulus, a single stimulus is presented, followed by another period without stimulus. The PSTH that result from such a paradigm are easy to interpret (see Fig. 6.5.). Spike density functions in RSVP are more challenging to interpret because of the lack of spontaneous activity to compare the response with. Testing the response to a stimulus against this pre-stimulus activity underestimates the magnitude of the response. Selectivity of cells was therefore analysed (see below) by contrasting the responses to different stimuli.

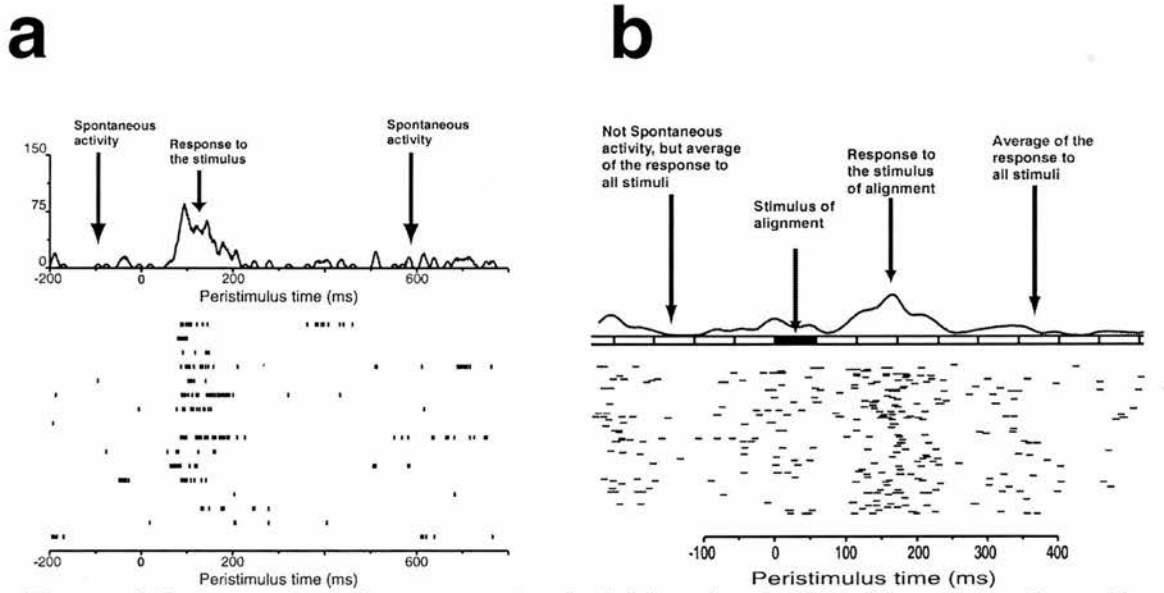


Figure 6.5: Comparison between a classical (a) and an RSVP (b) peristimulus spike density function and rastergram. (a) In a classic paradigm, a blank screen precedes the stimulus of alignment that is presented at t=0. Hence, activity prior to t=0 is spontaneous activity and the response to the stimulus is easy to associate with the stimulus, since it is the only complex stimulus being presented. (b) In RSVP, the situation is different. Stimuli are presented constantly (open boxes under the spike density function, presentation rate of 56ms/image in this case). The order of the stimuli is random, hence except for the stimulus of alignment, all other position will contain a stimulus chosen at random on each trial, possibly the stimulus of alignment itself. In interpreting an RSVP spike density function, one thus has to take into account that activity occurring before or after the stimulus of alignment is not spontaneous but stimulus caused activity. A "noisy" baseline hence does not have to mean that the neurone is unreliable: it may just be processing mostly other stimuli. Nevertheless, if the neurone responds to the stimulus of alignment with above or below average spike rates, the activity associated with that particular stimulus will

emerge above or below the rest of the activity at t=response latency. [(a) is adapted from Rolls et al., 1999; (b) is adapted from Keysers et al., 2000]

6.7.2. Latency detection

The responses of each neurone to each of the 8 stimuli of the test set in the 111ms/image testing condition were measured in a time window starting at 100ms post-stimulus onset and lasting 111ms. The stimuli eliciting the largest and smallest responses were defined as the neurone's 'best' and 'worst' stimuli. Response onset latency of a given neurone was computed off-line from trials for the best stimulus pooled across all presentation rates except the 14ms/image condition. The latency of response onset was defined as the first 1ms time bin at which the firing rate exceeded the mean + 2.58sd (i.e. $p<0.005$) of activity measured in a control period 250ms before stimulus onset, for at least 25 consecutive bins. Latency aligned responses refer to responses time-shifted by the difference between an individual neurone's response onset latency and the population average.

For experiment 3, investigating the effect of introducing gaps, latency was measured separately for conditions with and without gaps.

Pooling trials from different presentation rates before measuring response onset is based on the assumption that latency is independent of presentation rate. It appears, that presentation rate may have some impact on latency, but the effect is small, and in the order of no more than 10ms, legitimising the pooling of all trials to increase the signal to noise ratio.

Other techniques have been used to determine latency and yielded comparable results. Two such methods were based on variance.

The first of these measured the variance between the spike density function of the 8 stimuli as a function of time. Before onset of the systematic response, the baseline to all 8 stimuli is similar, and the variance low. During the systematic response, the spike density function will raise for effective stimuli and decrease for ineffective stimuli, causing the between stimulus variance to increase. After the end of the response, all 8 spike density functions will again be rather identical. Latency was then defined at as the moment in which the between stimulus variance rose above the confidence interval of the variance in the interval –250 to 0ms relative to stimulus onset.

The second method investigates variance between trials to each stimulus. For any given stimulus, spike rates should be similar during the systematic response, and dissimilar during the time before and after, when individual trials contain different stimuli. Latency was defined as the point at which between trial variance decreased below the confidence interval of the –250 to 0ms period.

6.7.3. Response duration

Response duration was assessed at the level of the population, and defined as the time during which the responses were different for the different stimuli. For each neurone, and presentation rate, a latency aligned spike density function was calculated using Gaussian filtering (with σ=5ms) for the best stimulus, and separately an equivalent function was calculated for the 'rest' (remaining 7 stimuli together). For each presentation rate, at the population level, the spike density functions for best and rest were compared using one entry for each neurone in a sliding matched pair *t*-test performed separately for each ms. No probability was computed when the response to the best stimulus was less than the response to the rest. Discrimination onset and offset were defined as the first ms where 30 consecutive 1ms bins have *t*-test $p<0.05$ and $p>0.05$, respectively. The average discrimination onset detected across presentation rate is 108.3ms (also the time of latency alignment) and is taken as the onset for the 'time window for response analysis' for all test rates. Note that the p values are not bonferroni corrected (see Appendix 3)

This population response duration does not necessarily reflect the response duration of an individual cell, but rather a type of average discrimination duration for all cells.

6.7.4. Uncontaminated responses.

As illustrated in Figure 6.5, in RSVP sequences effective stimuli can occur throughout the time surrounding a particular stimulus. The duration of a response to a stimulus was found to be longer than the duration of the stimulus itself by ~60ms. Hence, if the response to stimulus X is analysed, the neuronal activity of neighbouring effective stimuli may last into the time window of response analysis of X, and hence 'contaminate' the response to X with its own response. For some analysis, it is crucial to avoid such contamination by excluding trials in which the best stimulus occurred

close enough to X to risk contaminating the response estimate of X. Response to a stimulus X is considered 'uncontaminated' if it was flanked by a sufficient number of consecutive 'rest' stimuli (R, i.e. stimuli other than the best stimulus) to ensure that responses to nearby best stimuli did not contaminate the window of analysis for X. Contamination arises because response duration exceeds stimulus duration. The exact criterion therefore depended on presentation rate: RRXRR (for 111-222ms), RRRXRRR (42-56ms), RRRXRRRR (14-28ms).

Uncontaminated responses are only used when explicitly indicated, namely in sections 7.2.5., Chapter 10, Sections 11.3.2.2. and 11.3.2.3. and Chapter 12. All other analysis are performed on all trials, irrespectively of the uncontamination criterion.

6.7.5. Quality of stimulus discrimination

The quality of neuronal responses was assessed using spike counts in a variety of intervals and using a variety of methods: ANOVA comparing the spike counts to different stimuli, neurometrics and information theoretical analysis. These specific methods will be explained in the relevant chapters.

7. Experiment 1: I. Single cells in STSa and RSVP without gaps

7.1. Introduction

Neurones in temporal cortex respond selectively to complex object such as faces (e.g. Bruce *et al.*, 1981; Perrett *et al.*, 1982; Desimone *et al.*, 1984; Rolls, 1984; Baylis *et al.*, 1985; Yamane *et al.*, 1988; Oram and Perrett, 1992; Perrett *et al.*, 1992). This property makes temporal cortex well suited for the investigation of object recognition. Indeed, unlike earlier visual areas, temporal cortex seems to follow the percept rather than the stimulus in binocular rivalry (Logothetis, 1998; see Chapter 2). The mechanisms underlying these complex responses are still poorly understood. The visual system is thought of as a chain of visual areas containing neurones with increasingly complex response properties, and temporal cortex, especially the cortex lining the anterior superior temporal sulcus (STSa) is thought to be at the top of this hierarchy. The concept of hierarchy is problematic considering the large amount of neural connections transmitting information from the "later" to the "earlier" visual areas (e.g. Hupé et al., 1998; Felleman and Van Essen, 1991, see in particular Bullier and Nowak, 1995, for a critical discussion of the arguments contradicting with the hierarchical model), but the increasing complexity of stimuli required to activate the neurone along the ventral stream (Kobatake and Tanaka, 1994) and the roughly increasing visual response latencies along the stream (e.g. Schmolesky et al., 1998) support the idea of a basic hierarchy in feed-forward visual processing, albeit with shortcut routes, where information can reach very high areas surprisingly quickly.

A number of investigators have gained valuable insights into the mechanisms of object recognition by placing the visual system under extreme time constraints. Thorpe and colleagues investigated how quickly a visual scene can be categorised as containing an animal. They showed that correct motor responses can be generated as quickly as 235ms (humans) or 190ms (monkeys) after the scene was presented (Fabre-Thorpe et al., 1998). Event related potential (ERP) studies with humans during the same task indicate that the visual categorisation was performed 150ms after

stimulus presentation (Thorpe et al., 1996). The authors conclude that such short reaction times can only be the product of a feed-forward brain architecture in which each neural stage generates its output based on the very first spikes arriving from the preceding stage (Thorpe, 1990).

Another line of investigation placing temporal constraints on vision uses the backward masking paradigm (as described in Chapter 3) in which the time available to process one stimulus is limited by presenting a second stimulus (the mask) shortly after the first. Kovács et al. (1995) and Rolls et al. (1994a, 1994b, 1999) could show that presenting a masking stimulus shortly after a target stimulus reduces the amplitude and/or the duration of the response of a neurone selective for the target. While the responses became very small at their shortest stimulus onset asynchronies (SOA), the single cell responses did not quite disappear even at the fastest presentation rates. The authors focused mainly on how this reduction parallels the reduced detectability of the target in psychophysical tasks. Their finding are also very valuable for understanding the mechanisms involved in object vision. At their shortest SOA, very little time was left for each visual area to process and send out the result of its analysis to the next area before being interrupted by the masking stimulus. Cellular responses were typically shortened to ~60ms for the shortest SOA (20ms). The fact that the neurones still exhibited some stimulus selective responses under those time-constrained conditions is indicative of a brain architecture that can work even in very limited time intervals.

Unfortunately backward masking paradigms have relatively long gaps between trials and follow a 'nothing – stimulus – mask' sequence. The significant cellular discrimination performance under the most rapid presentation conditions is surprising but may thus be limited to this exceptional situation, in which during an interval without complex stimulus, a warning tone and a fixation dot are telling the visual system "Beware! A stimulus is going to appear soon!". Under normal viewing conditions, a visual event is usually both preceded and followed by other events, and hence both forward and backward masking are often combined. Uttal (1969a,b) showed that the effect of combining forward and backward masking cannot be predicted by simply adding the effects of forward and backward masking measured alone. To understand visual shape processing, it is thus important to measure the effect of combining forward and backward pattern masking on single cells in areas directly involved in visual object processing, such as IT or STSa.

In RSVP images are presented sequentially and continuously, with each image replacing the previous at the same location on the screen, one after another. RSVP thus combines multiple forward and backward masking, and gives us the tool to investigate the impact of that combination on the object selective responses of single neurones in the higher visual cortex. While RSVP has been used extensively in behavioural investigations (e.g. Potter & Levy, 1969; Chun and Potter, 1995; Subramaniam et al., 2000; and Chapter 5), it has so far not been used in conjunction with single cell recordings in the higher visual cortex. The study reported in this chapter will thus investigate for the first time the performance of object (including face) selective neurones in the STSa under continuous, RSVP conditions.

RSVP sequences composed of colour photographs of faces, everyday objects familiar and unfamiliar to the subjects, and naturalistic images taken from image archives were used as stimuli and presented in the centre of a computer screen. Neurones that responded selectively to complex patterns (e.g. faces, Oram and Perrett 1992) were recorded in the STSa while the monkey fixated such RSVP sequences for fruit juice reward. Each neurone was initially tested with up to 60 stimuli presented in random order at a moderate presentation rate (111ms/image) to determine effective stimuli for that particular cell. Where reliable responses were found, 8 stimuli were selected (2 best, 2 worst, 4 intermediate at driving the neurone out of the 60). Image sequences were then presented as permutations of these 8 stimuli shown successively without inter-stimulus gaps at durations ranging from 14 to 222ms/image (Table 7.1, see Chapter 6 for detailed methodological descriptions). The goal of the study was to investigate how the response of object selective neurones at the 'top' of the visual cortical hierarchy would deal with such extreme time constraints.

Label	Stimulus [Frames]	Gap [Frames]	Stimulus duration [ms]	ISI [ms]	SOA [ms]	Cells tested
14ms/image*	1		4.5	9.4	13.9	22
28ms/image	2		18.4	9.4	27.8	34
42ms/image*	3		32.3	9.4	41.7	23
56ms/image	4		46.1	9.4	55.6	34
111ms/image	8		101.7	9.4	111.1	34
222ms/image	16		212.8	9.4	222.2	34
	1	3	4.5	51.1	55.6	21
	2	6	18.4	92.8	111.1	34

Table 7.1: Conditions used in Experiment 1. For each condition, the label that will be used in this chapter is given together with the numbers of frames of stimulus used on the 72Hz monitor, the actual stimulus duration given the phosphor decay time of the screen, the interstimulus interval (ISI) between end of one stimulus and beginning of the next, the stimulus onset asynchrony (SOA) separating the onset of one stimulus from the onset of the next stimulus, and the number of neurones tested. See Section 6.5.1 for how these values have been obtained. Note that the two last conditions in the table, overlaid in grey, have been collected together with the other conditions, but are part of experiment 3, and are not going to be discussed in the present chapter.

7.2. Results

137 neurones were tested in two monkeys. 103/137 neurones were not fully tested, because none of the 60 stimuli tested evoked robust responses (~70% of cases) or because an effective stimulus was found, but the cell could not be recorded for long enough to complete testing (~30%). Thirty-four neurones were recorded long enough to complete testing at presentation rates of 222, 111, 56 and 28 ms/image and 23 of these neurones were tested additionally at 42 and 22 cells at 14ms/image. All of these neurones have been histologically attributed to the upper bank, the lower bank or the fundus of the STSa.

7.2.1. Qualitative discussion of an example

Figure 7.1 depicts the aligned responses of one neurone to the occurrences of face and half profile views of a monkey in the sequence. These stimuli caused the largest and second largest response, respectively, from those tested for the neurone in the 111ms/image reference condition (and hence are defined as the 'best' and 'second best' stimulus for the neurone).

This neurone was sharply tuned to the best stimulus both within the 30 stimuli of the search and the 8 stimuli of the test set (see chapter 6.6.2. for definitions of these terms): its response to the second-best stimulus was barely detectable, while its response to the best stimulus is strong and reliable. Other neurones had broader tuning.

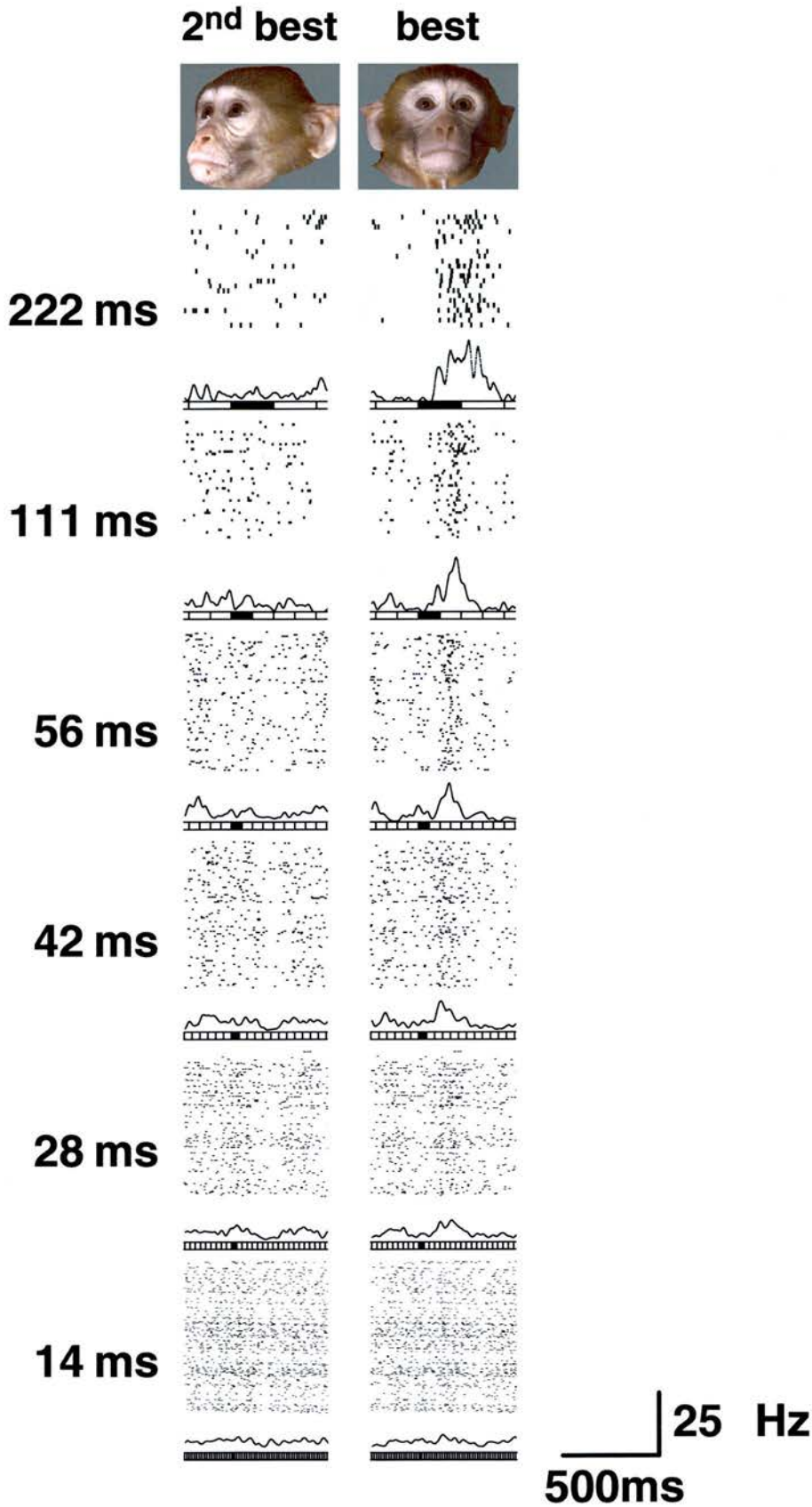


Figure 7.1: Responses of one neurone to the best (face) and second best (half profile) images at the different presentation rates. For each rate, rastergrams record activity during image sequences aligned for each occurrence of the face or profile images. Spike density function (smoothed using Gaussian filtering with s=10ms) combine the activity across all occurrences of the image (n= 24, 48, 96, 120, 192, 384 trials for rates 222-14ms/image). The black horizontal rectangles represent the time of presentation of the stimulus of alignment, the unfilled rectangles the timing of the other randomised stimuli in the sequence.

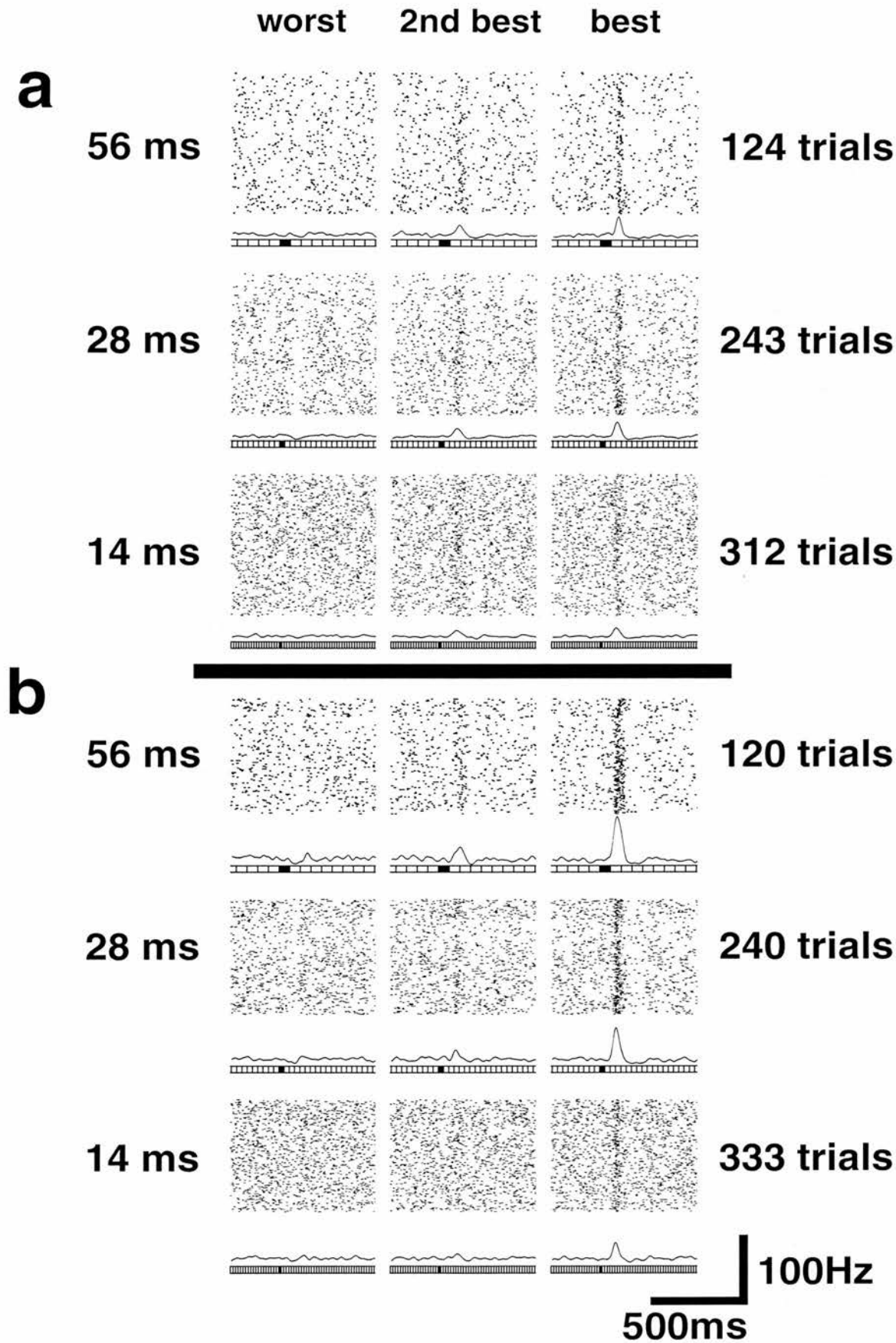


Figure 7.2: Responses of two cells at different presentation rates of 30 stimuli. Using the same conventions as Fig. 7.1, the responses at presentation rates of 14, 28 and 56ms/image are displayed for the best, second best and worst stimulus. These two cells were tested with 30 rather than 8 stimuli, resulting in a more pronounced response and a less elevated average firing rate in the rapid presentation conditions due to the less frequent appearance of the most effective stimulus. All spike density functions have the same scale indicated in the bottom right corner of the figure.

The responses of this neurone occurred at a relatively constant time interval (~90ms=response onset latency) after onset of the best face stimulus. The response lasted longer than the stimulus itself. This can be seen in Figure 7.1 by comparing the length of the solid black boxes under the spike density function (sdf) with the duration of the deflection in the sdf. Responses are evident on the majority of trials, especially for rates of 222 to 42ms/image. A stimulus-differentiating response is apparent, even at the faster rates (28-14ms), as a small peak in the spike density function for the best stimulus compared to that for the second best stimulus. The neurone illustrated in Fig. 7.1. and others (e.g. Fig. 7.2) responded in a consistent manner, time-locked to the onset of the best stimulus.

As discussed in section 6.7.1, the classical concept of a pre-stimulus baseline is problematic in RSVP. In other paradigms such as masking, the activity occurring before the stimulus presentation reflects the spontaneous activity level of the neurone (Fig 6.5). In our RSVP testing, because each stimulus is preceded by other randomly selected stimuli, the activity occurring before a stimulus of alignment is not spontaneous activity but the average response to all 8 stimuli used in the sequence (including in $1/8^{th}$ of cases, the best stimulus for the cell). A prestimulus period of 250ms will for example contain at least 2 occurrences of the best stimulus in each trial in the 14ms/image condition, resulting in an increased average firing rate compared to slower presentation rates. The faster the presentation rate, the more the baseline will be elevated. This is clearly visible in Fig. 7.1. Testing the response to a stimulus against this pre-stimulus activity would thus underestimate the magnitude of the response. Selectivity of cells was therefore analysed (see below) by contrasting the responses to different stimuli. Despite this elevation in average firing rate, at 14ms/image the neurone illustrated in Fig. 7.1 did not show the flat spike density function that would have occurred if presentation at this extremely high rate had caused the 8 overlapping images to be integrated and fused into a single compound image, as may have been expected based on the ~60ms period of integration typical for the visual system described in chapter 3.

If more than 8 stimuli are used in RSVP, the responses to the most effective stimulus will contrast more against the activity in the period prior to this stimulus. In a separate set of experiments that are not part of this thesis, Edwards et al. (in prep) recorded 21 neurones with RSVP sequences involving 30 rather than 8 stimuli. While these cells will not be included elsewhere in this thesis, they show that using more

stimuli results in clearer responses after alignment because the effective stimuli occur less often in the period surrounding the stimulus of alignment (1/30 vs. 1/8). The responses of two such cells are shown in Fig. 7.2. They illustrate that a clear response can occur to 14ms stimuli embedded in a continuous sequence of 30 unrelated images.

When presentation duration was reduced from 222ms to 111ms/image, only the duration of the response seemed to change while peak firing rate was unaffected. When presentation duration was reduced further to 56ms/image, the peak firing rate started to be reduced as well as the response duration. For durations of 56 to 14ms/image, response duration was barely effected, while response magnitude clearly decreased.

7.2.2. Time course of the population responses as a function of presentation rate.

A pattern similar to Figs. 7.1 and 7.2 was evident when the neurones that had been tested individually were considered as a population (see Box 7.1). Response latency varied between neurones with a mean of 108ms (range 56-171ms, Fig. 7.3). Figs. 7.4a, b depict the average responses of all neurones to their best ('best') and to the other stimuli ('rest' = average of all 7 stimuli except the best) after the responses have been aligned to the same latency (108ms, see section 6.7.2.).

As for the single cell of Figure 7.1., reducing image duration from 222ms/image to 42 ms/image reduced response duration without much effect on peak firing rate. For shorter SOA, duration remained above ~60ms, but peak firing rate decreases (Fig. 7.5a). If the population spike density functions are aligned on stimulus offset rather than onset, a form of neural persistence becomes apparent. If response duration was equal to stimulus duration, responses should decrease sharply at stimulus offset time plus cell latency (Fig. 7.5b). This was not the case: baseline activity was not reached earlier than 60ms after this time, indicating a decaying neural iconic memory of ~60ms. This persistence is not due to the 10ms standard deviation Gaussian filter kernel used to produce the sdf (see Fig 7.5c).

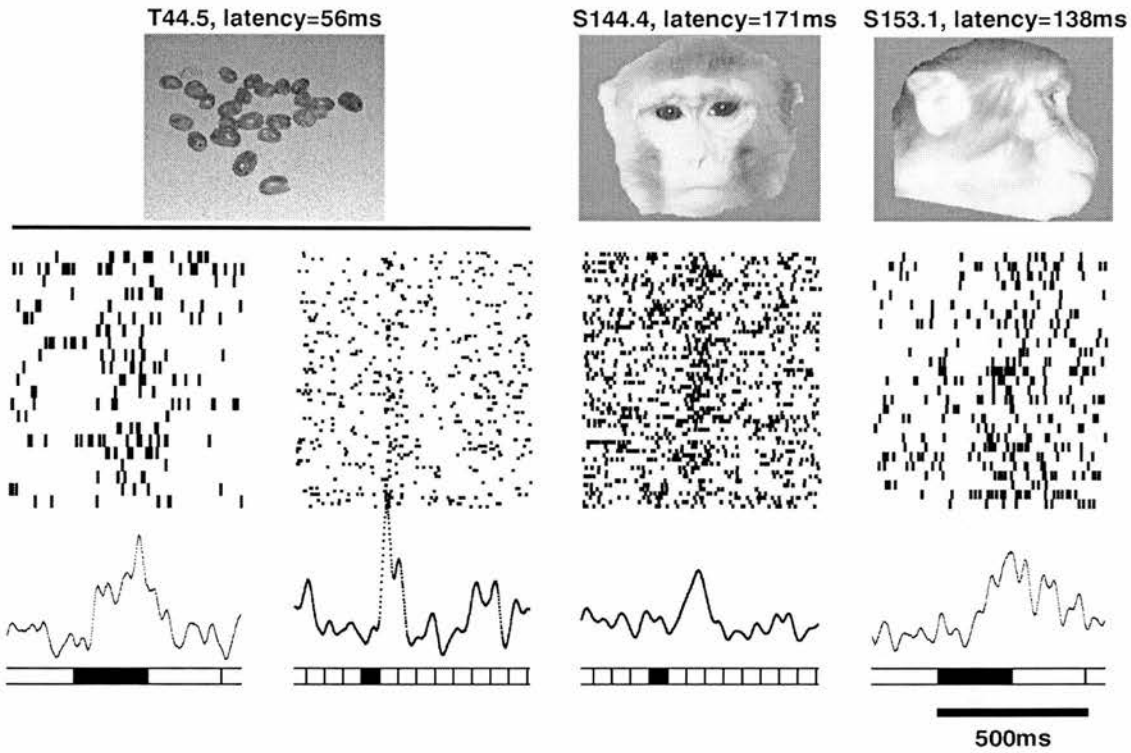


Figure 7.3: Rastergram, spike density functions, best stimulus and detected response latency (indicated over the best stimulus) for 3 cells in STSa illustrating the range of response latencies observed. From left to right: cell T44.5 at 222ms/image and 56ms/image; cell S144.4 at 56ms/image and cell S153.1 at 222ms/image.

Box 7.1: Population analysis

In the brain, more than one cell generally respond to a particular visual stimulus. The processing of a particular stimulus is then achieved by the population of neurones that respond to a particular stimulus. Electrophysiologically, it is till impossible to record simultaneously from all single cells responding to a particular stimulus simultaneously. Nevertheless, a 'trick' can be used to get a first approximation of such a population perspective of stimulus processing: although the different cells recorded in the present thesis often respond to different stimuli, one pretends that they all responded to the same 8 stimuli, called 'best' to 'worst' stimuli. This is done under the assumption, that if we had recorded from enough cells, we could have found 34 cells all responding to the same set of 8 stimuli, and that the way these hypothetical 34 cells respond to their common 8 stimuli is not systematically different from the way the 34 cells we actually recorded from responded to their individual sets of 8 stimuli. A similar approach has been used by Newsome and co-workers (e.g. Newsome et al., 1989) and more particularly in Oram and Perrett (1992).

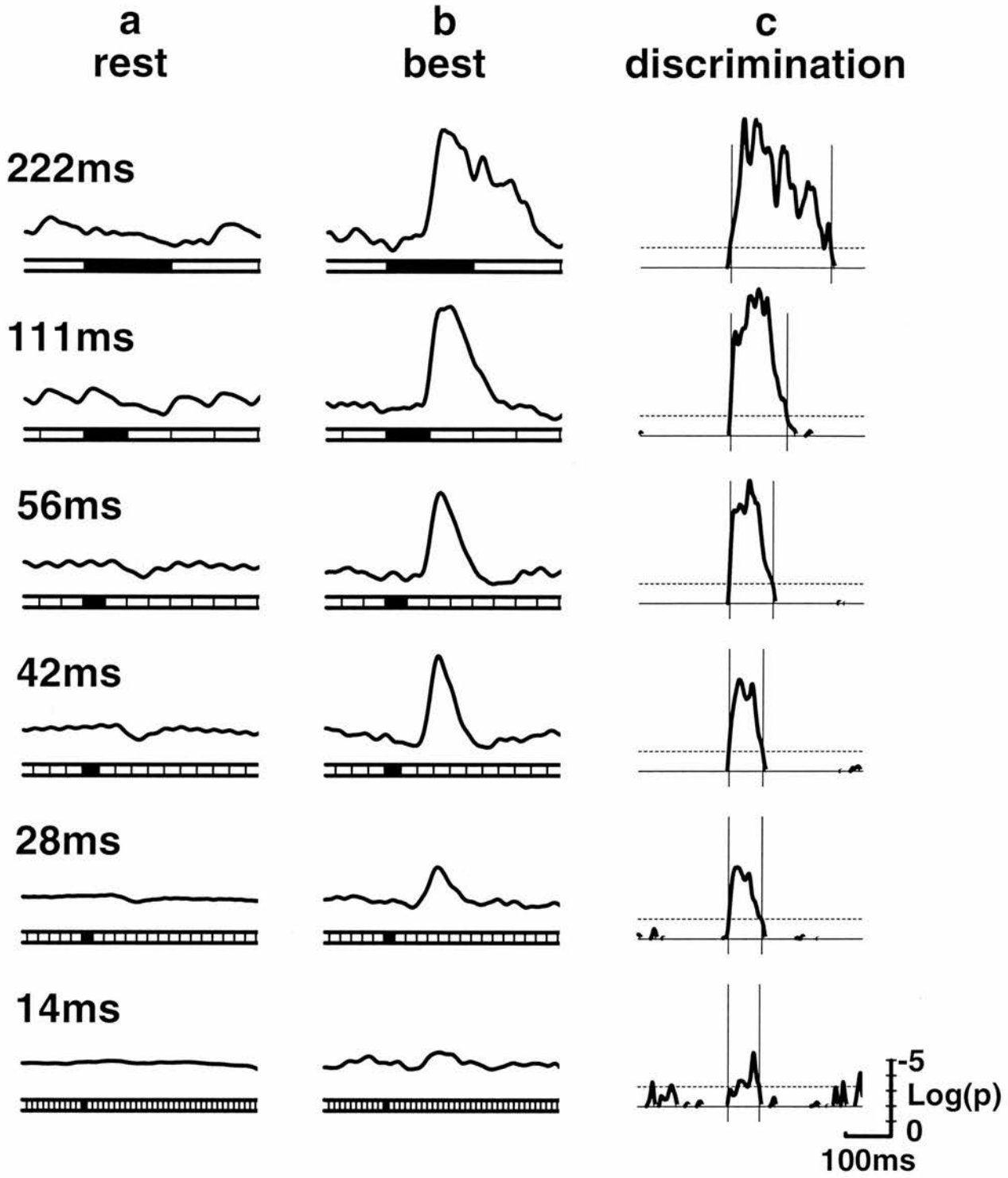


Figure 7.4: Normalised average neurone responses for effective and ineffective stimuli as a function of presentation rate. For each test rate, the responses to the best (b) and the rest (a) of the stimuli (defined at the reference test rate of 111ms/image) were latency aligned (see Chapter 6), normalised, averaged over neurones, and smoothed (Gaussian with s=10ms). See Table 7.1 for number of neurones per condition. The black horizontal rectangles represent the time of presentation of the stimulus of alignment, the unfilled rectangles the timing of the other randomised stimuli in the sequence. The vertical scale is identical for all presentation rates. (c) Probability of discrimination between stimuli as a function of time. For each neurone, and presentation rate, a latency aligned spike density function was calculated using a Gaussian filter (with s=5ms) for the best stimulus, and separately an equivalent function was calculated for the rest (remaining 7 stimuli together). For each presentation rate, at the population level, the spike density functions for 'best' and 'rest' were compared using one entry for each neurone in a sliding matched pair t-test performed separately for each ms. No probability was computed when the response to the best stimulus was less than the response to the rest. Discrimination onset and offset were defined as the first ms where 30 consecutive 1ms bins have t-test $p<0.05$ and $p>0.05$, respectively. The average discrimination onset detected across presentation rate is 108.3ms (also the time of latency alignment) and is taken as the onset for the 'time window for response analysis' for all test rates. The vertical lines represent the beginning and end of this window, the dashed horizontal line the $p=0.05$ criterion. Note that the p values are not Bonferroni corrected and are used only to determine the 'time window for response analysis' while the significance of the population response is assessed separately.

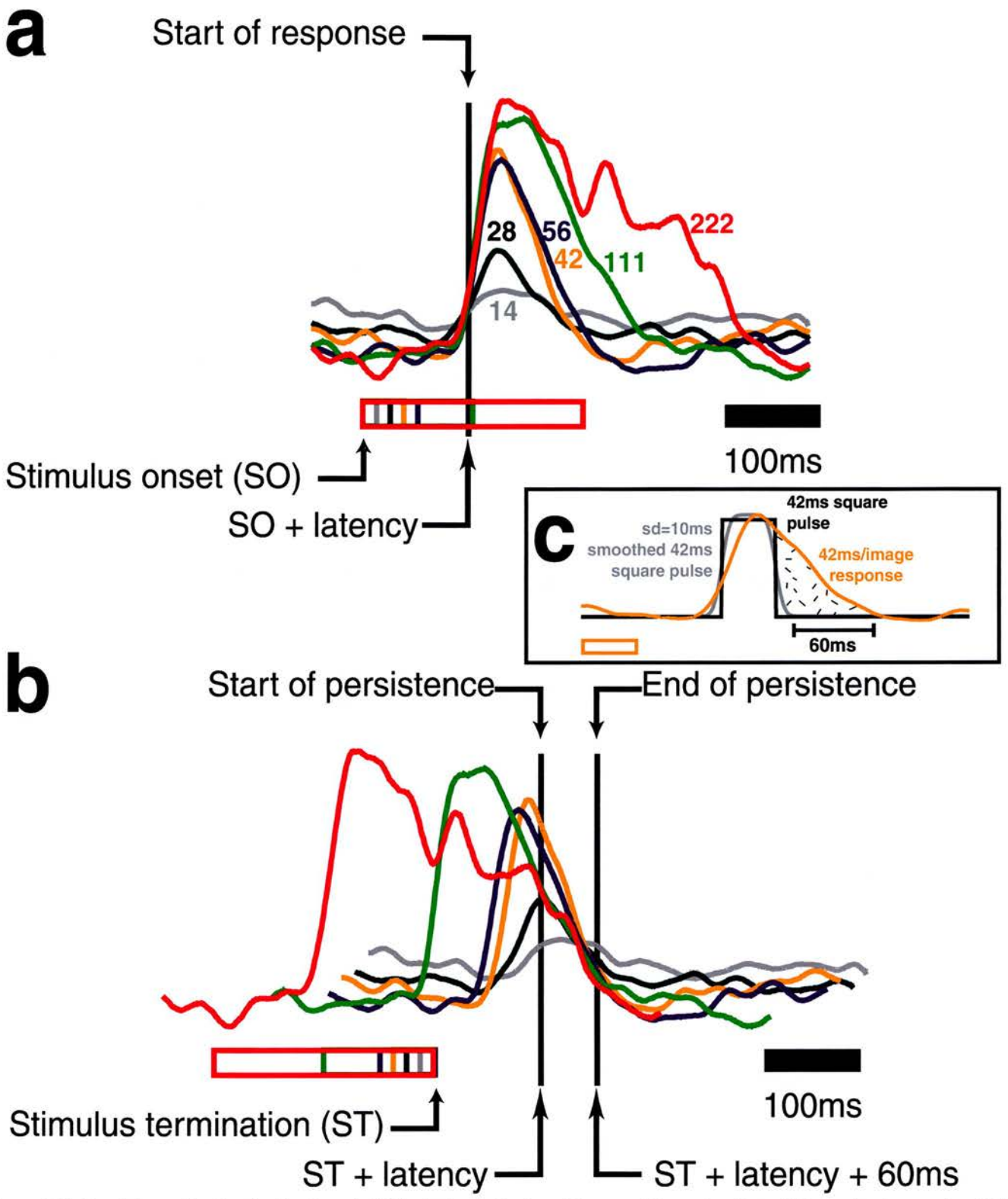


Figure 7.5: Population spike density functions (sdf) for the best stimulus at 6 presentation rates aligned (a) on stimulus onset and (b) on stimulus termination. The timing of the stimulus is indicated by an open box of the same colour as the sdf. Presentation rate is indicated in arabic letters next to the relevant sdf in (a) and is to be read in ms/image. The same colour-code is used in (b). **(a)** indicate how shortening the presentation duration reduces response duration but not peak firing rates for SOA>60ms, and peak firing rate but not response duration for SOA<60ms. Note how the responses start briskly at stimulus onset (SO) + population latency. **(b)** shows clearly the neural persistence of STSa cells. Without persistence, the sdf should briskly return to baseline at stimulus termination (ST) + latency. Instead, the responses stay elevated until 60ms later, showing a progressively decaying iconic memory of 60ms duration. **(c)** Illustrates that the persistence observed in (b) is not due to the 10ms gaussian kernel used to create the sdf. A square pulse (black) of 42ms is smoothed using a s=10ms gaussian kernel as in (a) and (b). The result in grey is compared with the population response to a 42ms stimulus (orange). Clearly, the hashed surface between the grey and the orange line is neural persistence that is not accounted for by the smoothing technique, and is lasting for about ~60ms beyond what is expected due to smoothing alone. Notice how the orange line reaches its peak only about 10-20ms after the smoothed square pulse, illustrating how close the neural response is to an instanteneous nothing-to-all response.

Despite the reduction in amplitude, a selective population response remains apparent certainly at 28ms/image, and to some extent even at the fastest presentation rate (14ms/image, Fig. 7.4)

The time course of stimulus discrimination was assessed by determining when the responses relate to the stimulus of alignment. To do this, a floating t-test was performed comparing at each millisecond the sdf for best and rest, using neurones as between subject and stimulus as within subject factor. The result is shown in Fig. 7.4c as the probability of the null hypothesis H_0:best≤rest, displayed as an inversed logarithmic plot. This plot can be read to mean that the higher the black curve, the more the firing rate of the cells discriminate between stimuli by firing more to the best stimulus.

Discrimination calculated in that way arises on average at 108ms and remains elevated for a duration that exceeded stimulus duration by ~60ms for all presentation rates (Table 7.2: columns 4, 5). This confirms the findings in Fig. 7.5b, and indicates that the neural persistence of ~60ms observed in the responses is stimulus specific, and is suitable to extract information about the stimulus.

7.2.3. The population of neurones discriminates at 14ms/image

The period of time of stimulus specific response determined in Fig. 7.4c was used as the "time window for response analysis". Spikes counted in that interval can be used to test whether the response discriminated between stimuli at different presentation rates.

ms/ image	images/s	Response onset*	Response Duration	Resp – Stimulus duration	Number of neurones	% neurones discriminating†
222	4.5	108	284	62	34	97
111	9		168	57	34	97
56	18		112	56	34	94
42	24		86	44	23	91
28	36		93	65	34	79
14	72		71	57	22	65

Table 7.2 Statistics of neurone performance as a function of presentation rate. The images were presented as a continuous sequence, so that the number of images per second is the inverse of the duration of each image. The duration of the population response was defined as shown in Fig. 7.4c. *Time of latency alignment and average detected discrimination onset. †: % neurones with ANOVA testing the effect of stimulus (1 to 8) on neurone response ($p<0.05$).

Following methods of Sary et al. (1993), the cell population's capacity to signal the presence of specific stimuli at fast rates was assessed. An average tuning curve (Fig. 7.6) was computed based on the activity of the population in the entire response duration starting at each cells response latency and lasting for the duration indicated in Table 7.2, column 4 ('Response Duration). On average, the stimulus that evoked the largest response at 111ms/image also produced the largest response at the higher presentation rates. Indeed, the rank order of stimulus effectiveness remains relatively stable across presentation rate although it becomes distinctively flatter at the most rapid rates. At all presentation rates, the surface under the sdf in the time window identified in Figure 7.4c and Table 7.2 (column 3 and 4) is significantly larger for the best stimulus compared to the rest (t-test with stimulus as within subject and neurones as between subject factor, $p<0.05$).

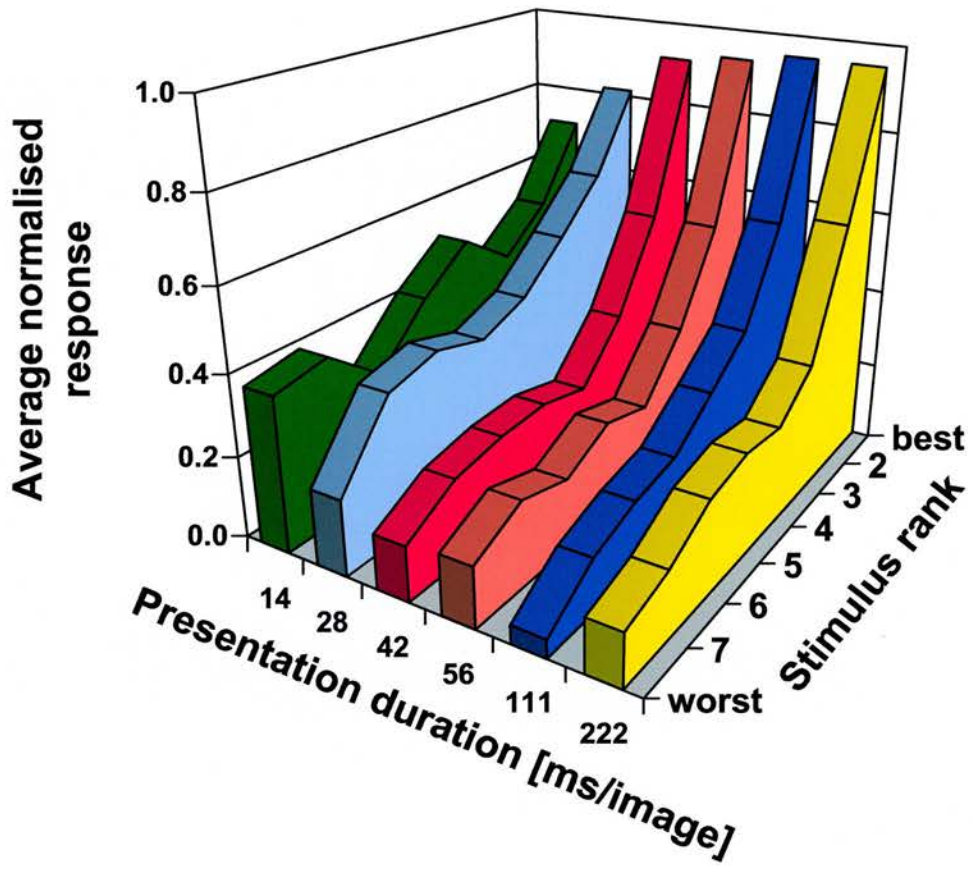


Figure 7.6: Response as a function of presentation speed and stimulus rank order. The stimulus ranking for each neurone was defined in the 111ms/image condition. The neurones' responses to these same stimuli were assessed from the number of spikes occurring in the response windows starting at the neurones' latencies and lasting the durations defined in Table 7.2, column 4. These responses were normalised per neurone to range from 0 (smallest spike-count for that cell at that speed) to 1 (highest spike-count for that cell at that speed) and then averaged across available neurones (Sary et al., 1993). Deviations from 1 for the best and from 0 for the worst stimulus reflect the fact, that for some neurones, at that presentation rate, the best stimulus was not the one having the largest spike count. Note that stimulus ranking is defined based on the 111ms/image presentation rate, and spike counts calculated in an arbitrary window lasting from 100ms to 211ms, while the normalised spike-counts in this figure are taken in 'time window for response analysis' starting at the cells individual latency. This difference in spike-counting window sometimes results in the second-best stimulus of the arbitrary window to be the best in the latency-aligned window.

7.2.4. 65% of the individual cells discriminate significantly between visual stimuli at 14ms/image

To examine the extent to which single neurones exhibit a preserved stimulus coding capacity, neurone by neurone statistical analyses were performed for each occurrence of the stimuli, comparing the response to the different stimuli with one spike-count entry for each stimulus occurrence. ANOVA performed on the spike counts of each neurone during the response duration window (starting at the cells latency and lasting the duration shown in Table 7.2 column 4) indicated a significant ($p<0.05$, two tailed) stimulus discrimination for the majority of the neurones at each testing rate (Table 7.2 rightmost column): 65% showed significant stimulus discrimination at the highest rate of 14ms/image.

7.2.5. Rapid emergence of discrimination

In Figure 7.4c it appears that discrimination as assessed with a floating t-test emerges suddenly, and produces a steep and sudden increase in discrimination at response onset. The population spike density functions (see Fig. 7.5a) also start to rise steeply at stimulus onset time + latency. Indeed, if compared to a smoothed square wave function (Fig. 7.5c), the neural responses reach their peak firing rate only 10-20ms later, indicating how close they are to switching instantaneously from no firing to maximal firing. The average point at which the t-test becomes significant is 108.3ms, which is equal to the time of latency alignment (108ms), indicating that the cells start discriminating as soon as they start responding.

This rapid emergence is not a by-product of latency alignment. Indeed it is also present in many single neurones: for each neurone, stimulus discrimination was tested immediately before and after the neurone's response onset latency. Half the neurones (17/34) discriminated between stimuli in the 20ms interval after but not before the onset latency. This was assessed by pooling all presentation rates for each neurone, counting the 'uncontaminated' spikes (see section 6.7.4.) for best and rest stimuli in the 20ms before and after the neurone's latency, and comparing best vs. rest stimulus in each time interval using a within subject *t*-test with a significance criterion of alpha=0.05. For 10ms intervals before and after, 38% (13/34) had this property. Thirty percent of neurones showed rapidly emerging discrimination even in the 14ms/image condition alone (20ms analysis). Stimulus discrimination can thus arise within 10-20ms of response onset.

7.3. Discussion

7.3.1. Preserved coding capacity at 14ms/image

The main finding of this experiment is that if a stimulus is most effective at stimulating a neurone at a slow presentation rate, this stimulus will remain effective at a more rapid presentations rate. Both in the population, and in a majority of single cells, this is true even at the most rapid presentation rate tested in this experiment (14ms/image). The fact that the same stimulus remains the most effective at a higher presentation rate is of cardinal importance. If a neurone is thought to participate in representing the stimuli it fires most to, this representation can only be preserved at higher presentation rates if at those rates the neurone continues to fire most to the same neurones.

7.3.2. Multitasking in the visual system

Single cells can discriminate stimuli at 14ms/image presentation duration. This presentation duration is very short, compared to the average latency of the neurones that were recorded in the STSa: 108ms. This creates a peculiar situation: by the time a neurone responded to one image (called 'target image'), 7 more images have entered the retina, as demonstrated in Equation 7.1.

$$\frac{latency}{image\,duration} = \frac{108ms}{14ms/image} = 7.8\ images \quad \text{[Equation 7.1]}$$

What will happen to these other images? It may be, that none of them are processed until the processing of the target stimulus is completed. But this is highly unlikely: response duration at 14ms/image was limited to about 71ms (see Table 7.2), a time much shorter than when the same 14ms/image stimulus was not followed by an other stimulus (this will be shown in Chapter 11). Indeed, for a 222ms/image stimulus, few people will doubt, that each stimulus is processed. In that case, ~60ms were necessary for the next stimulus to interrupt the response to the target. Considering that ~60ms persistence was observed at all presentation rates (see Table 7.2, Figure 7.5b), it is likely that interruption always takes ~60ms to occur. Hence, to stop the response after 71ms in the 14ms/image condition, the interrupting stimulus must have occurred ~11ms (i.e. 71ms-60ms) after the target stimuli, and hence would be the stimulus directly following the target. Since stimuli are presented in random

order, the next stimulus is equally likely to be any of the 7 remaining stimuli, and hence we have some indirect evidence that all stimuli are generally processed well enough for their processing to interrupt the processing of the preceding stimulus.

We are presently planning to perform multiple cell recordings, in which 2 cells will be recorded, selective respectively for instance for a front and side view of a face, and present sequences in which a front and a side view follow each other in the sequence, to check if on a given trial, both are successfully processed. Until these results are available, the following conclusions can be proposed for the case were the multiple cell recordings indeed confirm what the response interruption suggest: that all 8 stimuli in the sequence are on average equally processed.

If all stimuli are processed equally, at a given point in time, a snapshot of the brain would look somewhat like Fig. 7.7. The details of the illustration are not important, and in particular, the 'serial' hierarchical model of visual processing illustrated in Fig. 7.7 has been rightly challenged in recent years (see for instance Bullier and Nowak, 1995; Schmolesky et al, 1998). What is important is that in RSVP, if all stimuli are indeed equally processed, the brain has to engage in 'multitasking', processing over 7 different stimuli at the same time. Multiple stimuli will be represented at the different levels of the visual system, and even within one area, due to the response duration outlasting stimulus duration and response latency differences between cells (range 56-171, see Fig. 7.3). This form of multitasking is different from the parallel processing of different aspects of a single stimulus (e.g. colour and motion), because it will include multiple, mutually exclusive representations of the same attribute, belonging to different images in the sequence (e.g. shape round and shape square). While the same is true every time a new object appears in a natural scene, in RSVP this situation is particularly emphasised.

As can be seen, in RSVP, if all stimuli are indeed processed equally, V4 will start to process one stimulus, when V2 will already start processing the next stimulus in the sequence. Under this situation, the many feed-back connections in the brain, which have been shown by cooling experiments to improve visual processing under classical presentation conditions (e.g. Sandell and Schiller, 1982; Hupe at al., 1998) would run the risk of being detrimental to visual processing. Feed-forward information about a new stimulus may be combined with feed-back information from an older stimulus (see red arrow in Fig. 7.7), potentially leading to 'feed-back

integration' of the two stimuli. The fact that significant stimulus discrimination can be achieved despite the situation illustrated in Figure 7.7 constrains models of feedback contributions to vision. Feedback regarding the representation of one stimulus must avoid disturbing the representation of another stimulus.

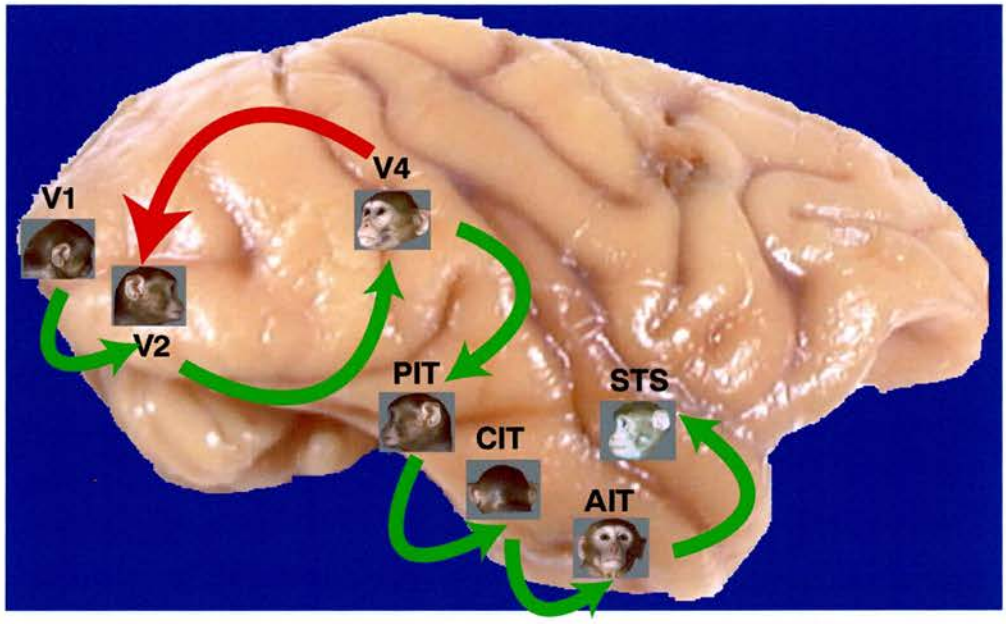


Figure 7.7: An impression of a snapshot of the brain activity during a 14ms/image RSVP sequence. In green, the forward connections of the ventral stream are shown schematically, with the name of the visual areas in black. The photos of monkey heads are images presented in that order in the RSVP sequence, and represent the leading edge of the neural activity representing that stimulus at that point in time. As can been seen, 7 stimuli are represented simultaneously, but with their leading edge at different levels in the hierarchy. Note in red, the example of a feedback connection from V4 to V2.

7.3.3. Latency differences and their influence on RSVP performance.

In the visual system, the magno- and parvocellular system have different processing speeds (Nowak et al., 1995), with the average parvocellular latencies in V1 and V2 being about 20ms later than the average magnocellular latency in the same areas. Both type of information converge onto V4 (Ferrera et al., 1994). For stimuli with long presentation rates, it is easy to conceive how the two types of information can converge: a magnocellular information will arrive first, followed later by the sustained parvocellular fine-grained detail. Indeed, in the temporal cortex, in some neurones, fine-grained information about face identity arrives sometimes later than more general information about a face as a general category of objects (Sugase et al.,

1999). But under the fastest presentation rate used in our experiment (14ms/image), the magnocellular input regarding a new stimulus would arrive at the same time as, or even earlier than, the parvocellular information from a preceding stimulus. It is unclear how the system is able to deal with such a situation. The inability to combine magno- and parvocellular information due to the time constraints of RSVP may be one of the reasons for the reduced response magnitude at the more rapid rates. To test this hypothesis, we are currently performing experiments in which the high- and low frequency components of an image are shown either together, or separately. If at the most rapid rates, magno- and parvo input are unable to fuse, then one would expect the response to a full-frequency image to be no larger than that to its magno or parvocellular response alone. Preliminary results indicate that such a simple model does not apply, but a full analysis of the data is still pending, and will not be part of this thesis.

What is extraordinary is how scattered visual latencies are throughout the brain, and how, especially in RSVP, where response duration is severly limited (about 71ms at 14ms/image), the brain is able to extract some information about the stimuli despite the scattering of latencies. Munk et al., 1995 and Nowak et al., 1995 measured latencies in V1 and V2 to flashed stimuli, and latencies in V1 ranged from about 30 to 120ms. The scatter (90ms) is thus larger than the duration of responses measured in our experiment (71ms). In STSa, we measured response latencies varying from 56ms to 171ms (see Fig. 7.3). It is remarkable, how the visual system is able to handle such discrepancies, in the context of rapid presentations. In particular, theories of visual consciousness sometimes postulate, that different aspects of a scene are perceived at different cortical levels (movement in MT, colour in V4, overall shape in IT), and that the overall, unified conscious perception arises through the integration of these different levels of perception (Zeki and Bartels, 1999). It is challenging for such theories to have situations with such large latency differences and such short response durations. It may be that as we reduce SOA in RSVP sequences, we make it increasingly difficult for the system to integrate information from sources having different latencies. The decrease in response magnitude as SOA is decreased below 60ms may be due in part to this restriction.

7.3.4. Response magnitude and response duration

The meta-analysis of the single cell recordings under masking condition presented in Chapter 3 (Kovács et al., 1995; Rolls et al., 1994a, 1994b, 1999; see Figs. 3.16, 3.18, 3.21) indicated that for pattern backward masking, as SOA is reduced, at SOA>60ms, the durations of the response is reduced but peak firing rate is not, while at SOA<60ms, the response duration is kept constant at ~60ms while the peak firing rate decreases. This indicated that at SOA>60ms, a backward mask seems to interrupt the processing of the target, while at SOA<60ms, the mask limits the response duration to ~60ms, and some form of integration - occuring as early as the retina - probably reduces the response to the target image. In addition, some form of competition between the representation of the mask and the target (Chapter 4 and Figure 7.8 inset 'lateral competition') may contribute to reduce the responses further.

This meta-analytical result is fully confirmed by our empirical findings. Even in RSVP, where forward and backward masking are combined, there is something special about an SOA of ~60ms. Both in single cells (Figs. 7.1, 7.2.) and in the population (Fig. 7.4 and 7.5a), two different processes are visible for long and short SOA. When SOA was reduced from 222ms-56ms, response duration is reduced from 284ms to 112ms, but peak firing rate remains fairly unchanged. When SOA was reduced from 56-14ms/image, peak firing rate decreased dramatically, while response duration remained above 60ms at all times (see Table 7.2).

7.3.4.1. Neural persistence of ~60ms despite the presence of a mask

Indeed, in our findings, it appears that there might be an alternative way to describe the data. Rather than stating that at SOA<60ms, response duration remains constant, it appears in our data that the duration by which response duration outlasted stimulus duration is constant. Independently of presentation duration, responses lasted ~60ms longer than the stimulus (see Table 7.2 column 5 and Fig. 7.5b) yielding to Equation 7.2a or Equation 7.2b[7]

Response duration [ms] = *SOA* [ms] + 60ms **[Equation 7.2a]**

[7] As ISI was kept constant at 9.4ms (see Table 7.1) in this experiment, it is impossible to decide whether SOA or stimulus duration is the determining factor of the response duration. In Chapter 11, we will show that Equation 7.2a is the more general formulation applying to conditions with and without interstimulus gaps.

Response duration [ms] = *Stimulus duration* [ms] + 70ms **[Equation 7.2b]**

This effect is not due to smoothing the responses (see Fig. 7.5c). What is remarkable, is that the ~60ms persistence occurred independently of stimulus duration and despite the presence of a mask. As seen in Chapter 3, *visible* persistence typically shows an inverse duration effect, being longest for the shortest stimuli, while *information* persistence does not. Hence, the *neural* persistence we measured in STSa may relate directly to *information* but not *visible* persistence. In addition, the fact that neural persistence occurred despite the presence of a mask indicates that the masking process takes about 60ms to build up to sufficient strength to fully interrupt neural persistence. This new finding places new constrains on models of neural processes, and will have to be taken into account in models of persistence that presently assume a fairly instantaneous interruption of the representation of a target by a backward mask (e.g. Loftus and Irwin, 1998). Interestingly, in flash suppression too, a new stimulus takes ~60ms to inhibit the representation of an old stimulus and replace it with the representation of the new stimulus in IT (see Fig. 2.3).

It may be argued, that the neural persistence of the stimulus representation is due to the use of stimuli that are not very effective masks, but Rolls et al. (1994a,b, 1999) used masks that prevented identification of a face if presented simultaneously with the face, indicating that they were strong masks, and nevertheless the response duration in their studies outlasted stimulus duration. The same is true for Kovács et al.'s (1995) investigation of masking.

This neural persistence of about 60ms in STSa as measured in the population is similar to the neural persistence in the retinal ganglion cells (e.g. Levick and Sacks, 1970, see Fig. 3.3), and in Area 17 of the cat (Duysens et al., 1985, see Fig. 3.2) indicating that the neural persistence is likely to remain fairly constant throughout the ventral stream of the visual system.

It should be noted that the persistence calculated here is based on the latency-aligned population. Examination of individual cells revealed that the same relationship applies to most single cells, with some cells showing slightly larger, and others, like cell T44.5 illustrated in Figure 7.3, showing slightly shorter persistence.

Neural persistence may also be important for learning: visual input that often occur one after another will create slightly overlapping responses, and can thus be associated with each other by hebbian synapses. If in the future, part of the succession

is seen, these synapses can then prime the cascade of associated neurones, which are likely to be activated shortly thereafter, and hence improve the performance of the system by creating a predictive neural network.

7.3.4.2. Interruption

A neural persistence of ~60ms independent of SOA indicates that the response to a stimulus is indeed interrupted by the following stimulus. Otherwise response duration would not follow SOA in this clear relationship: response duration = SOA + 60ms. As outlined above, interruption takes ~60ms to occur, and appears in the population to occur in a progressive way over these 60ms (see Fig. 7.5).

Why and how this interruption occurs cannot be concluded from our data. One possibility is that the representation of objects might compete against each other (as described in Chapter 5), resulting in the representation of the new image inhibiting the representation of the preceding image, and eventually taking over, as the preceding image receives no more direct retinal input.

7.3.4.3. Integration and response amplitude

Neural persistence of ~60ms throughout the visual system despite the presence of a mask has strong consequences for the neural processing of RSVP sequences (see Figure 7.8). Due to neural persistence, the representation of the different stimuli in an RSVP sequence will overlap in time within a single cortical area. While the definitive proof of this overlap would necessitate simultaneous recording of at least two cells responding to two stimuli that follow each other in the sequence, the data reviewed in chapter 3 indicates that persistence exists throughout the visual system, as early as the retina, where attentional effects are unlikely, supporting the idea of an automatic phenomenon, that will occur equally for all stimuli in the sequence. When the representation of more than one stimulus occurs simultaneously in the brain, there are two ways in which this can reduce the amplitude of the response to the stimuli in the sequence (Fig 7.8):

'Feed-forward integration' can be best understood by an example. Imagine a sequence that contains a back view of a face followed by a front view. A postsynaptic neurone responding to a front view of the face will possibly receive excitatory input from presynaptic neurones responding to the features of the nose, the mouth and the eyes contained in the front view, but also possibly inhibitory inputs from neurones

representing the hair, the back-view of the ears contained in the persistent representation of the back-view that is no longer on the screen. This inhibiting information will reduce the post-synaptic activity, resulting in an overall activity that probably resembling that to the presentation of an overlaid front- and back-view of a face.

In '**lateral competition**', circuits between neurones representing the front and back views of the head may inhibit each other (see chapter 4), mutually reducing each other's responses.

Both mechanisms can work together, and would work at all levels of the visual system. For SOA<60ms, as SOA is reduced, more and more stimuli will fall within the ~60ms integration window created by persistence, reducing more and more the response to the stimulus (Fig 7.8), and the entire response will be affected by integration with the preceeding and/or following stimulus. These two mechanisms (lateral competition and forward integration) can thus explain why the amplitude of the response decreases as SOA decreases for SOA<60ms.

Both lateral competition and forward integration may be supplemented in their amplitude reducing effect by the feedback integration described in Figure 7.7. The large latency differences within the visual system (section 7.3.3) may create even more potential for erroneous integration of information belonging to neighbouring stimuli in the sequence, especially as SOA becomes increasingly short.

For longer SOA (>60ms) the situation is different. At the time of peak firing of a given neuron (e.g. the black neuron in Fig. 7.8), there will be always exactly one stimulus being represented at the same time, namely the directly preceding one, and hence, peak firing rate should not depend on SOA, which is exactly what we found. Integration with the preceding stimulus can only influence the first 60ms of the response and integration with the following stimulus can only affect the 60ms of persistence of the stimulus. For very long SOA (e.g. 222ms), a substantial part of the response will therefore be completely unaffected by integration, which may explain why at longer SOA, stimuli are perceived as non-integrated. Latency differences and

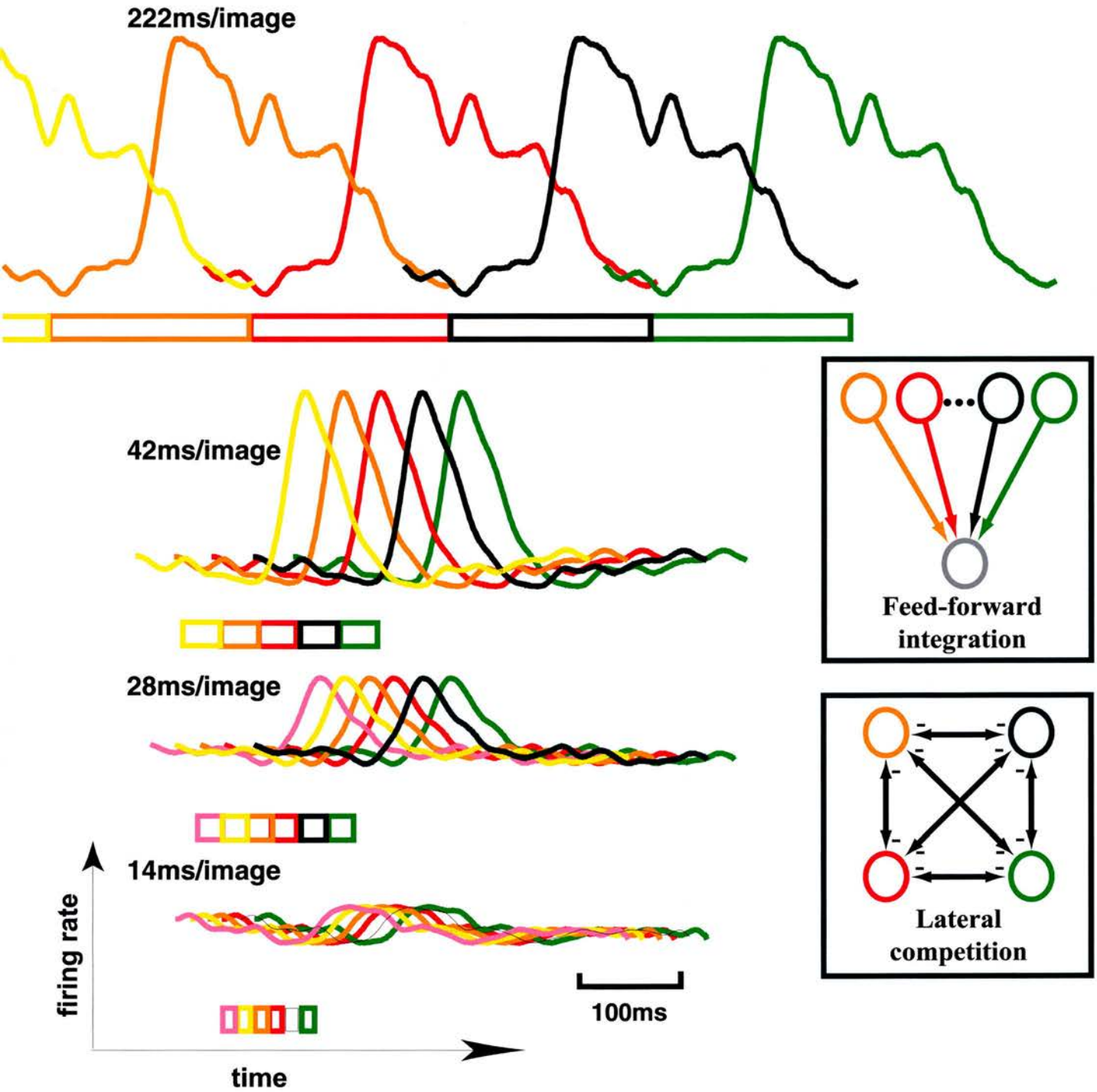


Figure 7.8: Consequences of ~60ms neural persistence during RSVP. A number of hypothetical neurones are represented at 4 presentation rates. Each neurone (circle in the insets) responds to a given colour-coded stimulus in the sequence, represented by the coloured box under the sdf. The sdf are not recordings from real neurones, but are copies of the population sdf calculated at that presentation rate and represent "average" responses. The representation of any particular stimulus (black) will be effected by the temporally overlapping representations of the previous (red, orange,yellow, pink) and following (green) stimuli. This temporal overlap can reduce response magnitude in two ways (see insets). Through **feed-forward integration:** a hypothetical postsynaptic neurone (grey) receives input from all these neurones and has no way to decide which activity corresponds to the persistence and which to the presence of a stimulus. Hence the grey neurone will have to treat the inputs of all presynaptic neurones equally, as if all stimuli were integrated on the screen. In a randomly ordered sequence this will decrease the signal to noise ratio of the black stimulus and hence decrease response magnitude in the grey neurone. The signal to noise ratio at any point in time is given by the height of the black curve divided by the sum of the heights of all other curves. Clearly, as presentation duration decreases, more and more stimuli will get within 60ms of the black stimulus, temporally overlapping with it and decreasing its signal to noise ration. A second mechanism is **lateral competition**, in which the neurones at one level of the hierarchy will mutually inhibit each other. In that case, the more cells are active during the firing of the black neurone, the more the black neurone will be inhibited. Both mechanisms probably occur simultaneously in the visual system. Note that in feed-forward integration, the interaction occurs without time-offset between the presynaptic neurones, while in lateral competition, at least one synaptic delay (~10ms) is necessary for competition to have an effect. Both forms of integration occur with preceding stimuli and subsequent stimuli.

feed-back integration may complicate this picture, but their effect is probably less pronounces, considering the quality of the responses at those longer SOA.

7.3.5. The issue of the reduced response magnitude and the number of trials required at high presentation rate

The magnitude of the response to the most effective stimulus at short presentation duration is no larger than that to a modestly effective stimulus at a longer presentation duration. It might be argued that therefore the brain cannot decode whether a given response is the response to a very effective but brief stimulus or a less effective, but longer stimulus. Yet at the population level, even if overall firing rates are changed, at any given time, a stimulus will activate more the cells that are tuned to it than the other cells. As long as cells still respond more to one stimulus than to the other, which at 14ms/image we have shown 65% of the cells to be capable of, the identity of a stimulus can thus still be decoded at the level of the population. A similar solution is used for extracting the colour of an object despite changes in illumination that will change the overall firing rate in the population.

In addition it might be argued that hundreds of trials are necessary to establish a significant difference between best and rest stimulus in a rapid presentation paradigm. The brain of course cannot wait to see an object 300 times before recognising it. But rather than observing 300 trials in one neurone, the brain can observe 300 neurones for one trial, which is similar (even though probably limited by common noise).

7.3.6. RSVP: a new method for testing stimuli

In addition to showing the preserved coding capacity at high presentation rate, the present findings have practical implications: they indicate that a neurophysiologist can present stimuli at very rapid presentation rate and nevertheless arrive to a stimulus tuning curve very similar to that that would have been obtained if a slower presentation rate had been used (Fig. 7.6) In our experiments, we equated the total presentation time at all presentation rate (see Chapter 6). That is to say, we collected twice as many trials for stimuli presented twice as short. Nevertheless it appears from the rastergrams (Figs. 7.1, 7.2), that at presentation times of 56ms/image or longer the large number of trials we collected was not necessary to recognise clearly the selectivity of the neurone. In addition, even if more trials need to be collected to

obtain a very clear response at higher presentation rates, this has the advantage of allowing a number of analyses (such as information theoretical analysis), which critically depend on a large number of trials. In the light of the present findings, it appears that rather than using conventional testing methods a neurophysiologist may use RSVP to collect either many more trials on the same number of stimuli, or use the same number of trials per stimuli and increase substantially the number of stimuli tested. This opens the prospect of confronting neurones with thousands of stimuli rather than the more restricted and potentially biased stimulus-sets used currently in visual neurophysiology. Investigating the feasibility of such rapid cell characterisation using RSVP was the original motivation of the present work.

7.3.7. Rapid onset of discrimination.

One of the interesting findings of the present chapter, is the rapid onset of stimulus discrimination. It appears, that within 10-20ms of response onset, the firing of a single cell is able to give significant information regarding the identity of a stimulus. This was shown both at the population level, as a steeply increasing discrimination curve in Fig. 7.4c, by the rapid occurrence of the peak firing rate (Fig. 7.5c) and at the single cell level. This result is similar to that found by Oram and Perrett (1992). In the next chapter, applying information theoretical analysis, the time course of information contained in the firing rate will be analysed in more details, and its implication for visual processing will be discussed.

7.4. Summary and Conclusions

Single cells in the STSa were recorded in two rhesus monkeys while RSVP sequences were presented to the subjects at presentation rates ranging from 222ms/image to 14ms/image. Both the population and a majority of single cells were able to discriminate between stimuli at the most rapid presentation rate (14ms/image=72images/s). This finding is surprising, considering that in the cinema projection of a motion picture, only 24 frames are presented each second (i.e. 42ms/image) and it is commonly assumed that a single frame (i.e. 42ms) would not be perceived. Indeed, at 42ms/image (identical to one motion picture frame) the cellular

response had almost the same peak firing rate in Experiment 1 as that to stimuli of longer duration. It should be noted, that for TV programmes the situation is slightly more complex than for the projection of a motion picture. Both the PAL system used in Europe, and the NTSC system used in the U.S.A. use an interleaved image presentation: each frame is divided in two half-frames. One half-frame contains all the odd lines, the other all the even lines of each frame. The two half-frames are presented one after the other to reduce flicker. Hence, while PAL presents 25 frames per second, it also presents 50 half-frames. NTSC presents 30 frames per second or 60 half-frames. It is therefore unclear if RSVP sequences should be compared to the half-frame or frame presentation rate of TV programmes. For cinema motion pictures this problem does not exist: the projector projects one frame, then a shutter prevents projection of the image while the projector moves to the next frame, and a new frame is projected and so on, at a rate of 24 frames per second.

The fact that at one of the very last stages of visual form recognition, the brain is able to successfully process unrelated images continuously presented at such rapid presentation rate, places strict constraints on models of visual processing. In addition, by showing that response duration = SOA + 60ms for RSVP sequences without interstimulus gaps, a new insight is given into the mechanisms underlying masking and RSVP. It appears that a new stimulus will indeed interruption the processing of a preceding stimulus at all presentation rates, but that unlike former findings suggested, interruption is not instantaneous, but takes about 60ms. Conversely, the present data demonstrate that cells in the STSa show a visual persistence of ~60ms beyond the onset of the directly following masking stimulus. This ~60ms persistence is similar to that found in the ganglion cells of the retina, and indicates that throughout the visual system a rather constant ~60ms persistence occurs when stimuli are directly followed by masks. Finally, for SOA<60ms, we observed that reducing SOA reduces peak firing rate. A number of mechanisms may participate in this response magnitude reduction, integration being almost certainly one of them, considering the known occurrence of integration in the early visual system. Altogether, the present data supports the notion of interruption occuring at all SOA, while integration, resulting in response magnitude reduction, occurs most prominently at SOA<60ms, and possibly also at longer presentation rates, but then only influencing part of the response.

7.5. Methods

Single cells were isolated, tested with 30-60 stimuli to identify 8 stimuli ranging from the best to the worst stimulus for that cell, and responses were then tested using these 8 stimuli at presentation durations of 222ms/image to 14ms/image. Cells were then tested individually or as a population after latency alignment. All methods are described in detail in Chapter 6, or the body of the text.

8. Information theoretical analysis

8.1. Introduction

The previous Chapter presented single cell responses to a set of 8 stimuli presented repeatedly using RSVP presentation at stimulus durations varying from 222ms/image to 14ms/image. It could be shown, that at all presentation rates, the population of neurones is able to significantly differentiate between stimuli, as assessed by 'classical' statistics, such as ANOVA.

Recently, information theoretical analysis has been increasingly applied to neural responses (e.g. Sugase et al., 1999; Gershon et al., 1998; Rolls et al., 1999). Brain areas concerned with the planning of actions have to base their decisions on the firing pattern of neurones in sensory areas. Hence, the function of the firing of neurones in sensory areas is to signal the presence of a particular stimulus. Information theory is a formal way to quantify how well the firing of a neurone signals what stimulus was present in the outside world and is thus in principle the method of choice for analysing neural responses.

One challenge in applying information theory to neural signals is choosing the right way to quantify the neural responses. Heller et al., 1995 demonstrated that ~80% of the total information contained in the response of a neurone is contained in the spike count of the response in a 300ms window. In the present chapter, only the spike count of the response will thus be measured, in the awareness, that information based on spike count underestimates total information by ~20%.

Describing the mathematical concepts underlying information theory would go beyond the scope of this thesis (see Jones, 1979 for an excellent mathematical introduction to information theory). Briefly, the mutual information *I(S,R)* between a set $S=\{s_1,\ldots,s_n\}$ of the n stimuli and a set $R=\{r_0,\ldots,r_l\}$ of l possible spike counts is calculated based on the equation (8.1), sometimes known as the Shannon Equation (Shannon, 1948).

$$I(S,R) = \sum_s \sum_r p(s,r) \log_2 \frac{p(s,r)}{p(s)p(r)} \quad (8.1)$$

where *I(S,R)* is the mutual information in bits of information, p(s,r) is the probability of stimulus s being associated with spikecount r, p(s) is the probability of stimulus s, and p(r) is the probability of response r.

Theoretically speaking, information is the measure of choice for analysing the data of Experiment 1. It yields a direct estimate not only of whether a particular stimulus causes a larger responses than other stimuli, but indicates how much the response of a neurone helps predict which stimulus was presented on the screen. The data gathered using all 8 stimuli used in Experiment 1 can be analysed using information theory with respect to how different the responses are that they produce in single STSa neurones. The more the responses caused by different stimuli will be different, the more a particular response will allow prediction of which stimulus produced it and the higher the mutual information between stimulus and response will be.

Practically speaking, information theory faces a serious problem: the problem of limited sampling. Equation 8.1 is correct, if the population probabilities of the system are known. *In vivo*, where sometimes only few trials can be collected for a given stimulus, knowledge of the population probabilities is sketchy at best. Directly using observed probabilities to calculate information using equation 8.1 consistently leads to overestimation of the true information contained in the responses (Optican et al., 1991; Chee-Orts and Optican, 1993; Treves and Panzeri, 1995; Panzeri and Treves, 1996), a problem known as the limited sampling bias. While correction factors for specific methods of information calculation exist (e.g. Panzeri and Treves, 1996), it remains best not to interpret information in terms of absolute values, but rather to interpret information in relative terms.

In the present chapter, the single cell recordings of Chapter 7 will be analysed using information theoretical analysis. Two aims will be pursued. First, measuring how reducing SOA in RSVP constrains how much information about the stimulus is contained in the response of single neurones in STSa. Second, measuring the time course of information in STSa responses during RSVP to identify *when* in time the following and preceding stimuli in the sequence affect the stimulus information available in STSa responses.

8.2. Method

For each cell, presentation rate and stimulus of alignment, spike count distributions were computed within a given time window. Such distributions indicate how often a certain spike count occurs after a particular stimulus (see Fig. 8.1).

In a second step, as a slight modification of methods of Földiák (1993) and Gershon (1998), the probability of a certain spike count 'r' given a stimulus 's' is estimated by the maximum likelihood fitting of a discreticised Gaussian, truncated below zero (i.e. $p(r,s)=0$ if $r<0$) (see Foldiak 1993), and normalised to sum to 1 (see Fig. 8.1). Such a probability density estimation using a truncated Gaussian has been shown to be a good model for spike count histograms of IT cells (Gerschon et al., 1998). This fitted Gaussian is then used instead of the observed conditional spike count frequencies to calculate the mutual information based on equation 8.1.

Sometimes in the spike count histograms, a high spike count occurred once, without the directly lower or higher spike count ever occurring. These 'odd' high spike counts tended to result in Gaussians that fitted all the rest of the data with less accuracy. For the purpose of this analysis, such single occurrences of high spike counts without the directly smaller and higher spike count ever occurring, were excluded from the data before the fitting procedure and considered outliers.

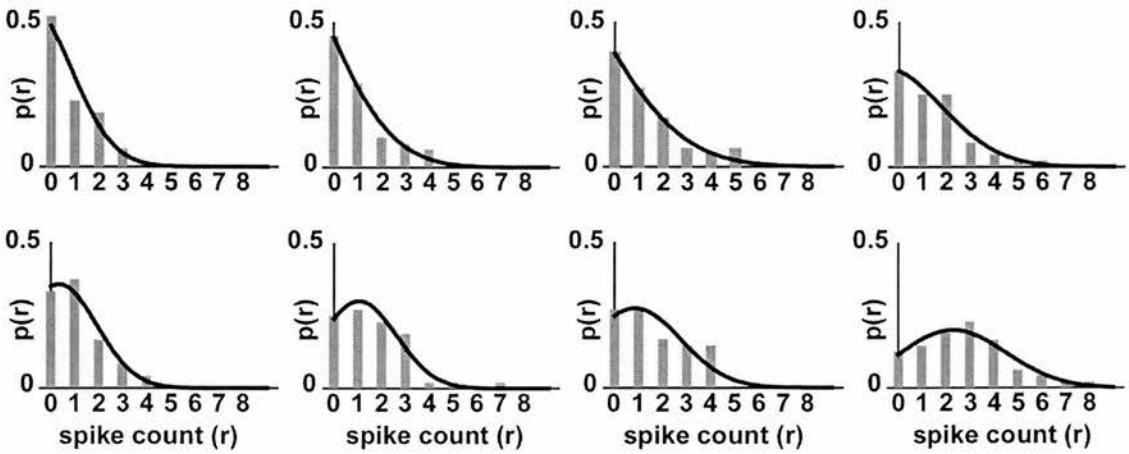


Figure 8.1.: Spike count histograms (grey bars) for 8 stimuli at 111ms/image for cell T44.1 (48 trials per stimulus) counted from the cell's response onset latency for 168ms. The stimulus on the bottom right was the best stimulus: high spike counts occurred frequently for that stimulus. The black curve represent the modified Gaussians fitted individually to each histogram based on a maximum likelihood procedure. For this cell and presentation rate, a Chi-Square comparing the observed probabilities with those predicted based on the fitted Gaussian indicated that the data

never significantly differed from the fitted Gaussians (df=8, all p>0.9), confirming that fitting a truncated Gaussian is a good description of spike count histograms.

8.3. Results and Discussion

8.3.1. *I(S,R)* contained in the overall response

First, mutual information (*I(S,R)*) was measured during variable spike count window covering the entire period of the response as defined in Fig. 7.4 and Table 7.2., i.e. for each cell, spikes were counted starting at the cells response onset latency and lasting for the duration indicated in Table 7.2 column 4. The results are indicated in Figure 8.2a as the black line. Clearly, reducing SOA reduced the *I(S,R)*.

To investigate if *I(S,R)* decreased as SOA was decreased because the spike count window is reduced, the analysis was repeated using the shortest, 71ms spike count window for all SOA. The results are indicated as the green line in Figure 8.2a. confirms that information decreases with decreasing SOA even if the spike count window is kept constant.

A repeated measurement 2 windows x 6 SOA ANOVA was performed on the *I(S,R)*measure. The ANOVA yielded no main effect for window ($F(1,13)=0.22$, $p>0.6$), indicating that counting spikes for longer than 71ms does not systematically increase information. There was a significant main effect of SOA ($F(5,65)=40$, $p<0.001$), indicating that reducing SOA reduces the *I(S,R)*. A Newman Keuls post-hoc analysis (critical range of 0.05) indicated that all SOA yielded *I(S,R)* values that differed from those at other SOA, except the 14ms/image vs. 28ms/image and the 42ms/image vs 56ms/image conditions which did not differ significantly ($p>0.05$). Finally, there was also a significant interaction between SOA and window ($F(5,65)=4.6$, $p<0.01$), and a Newman-Keuls test indicated that only for the 222ms/image and 111ms/image conditions was there a significant ($p<0.05$) difference between the *I(S,R)* contained in the two spike count windows. Results were identical if the effect of SOA was tested using non-parametric, one way Friedman ANOVA.

At shorter SOA, more trials were available to assess *I(S,R)* (Fig. 6.4). To investigate how trial numbers affected *I(S,R)*, and to assess the reliability of the assessment method, only 40 trials (or the maximum trial number available, whatever value was smaller), picked at random from the available trials, were used for each

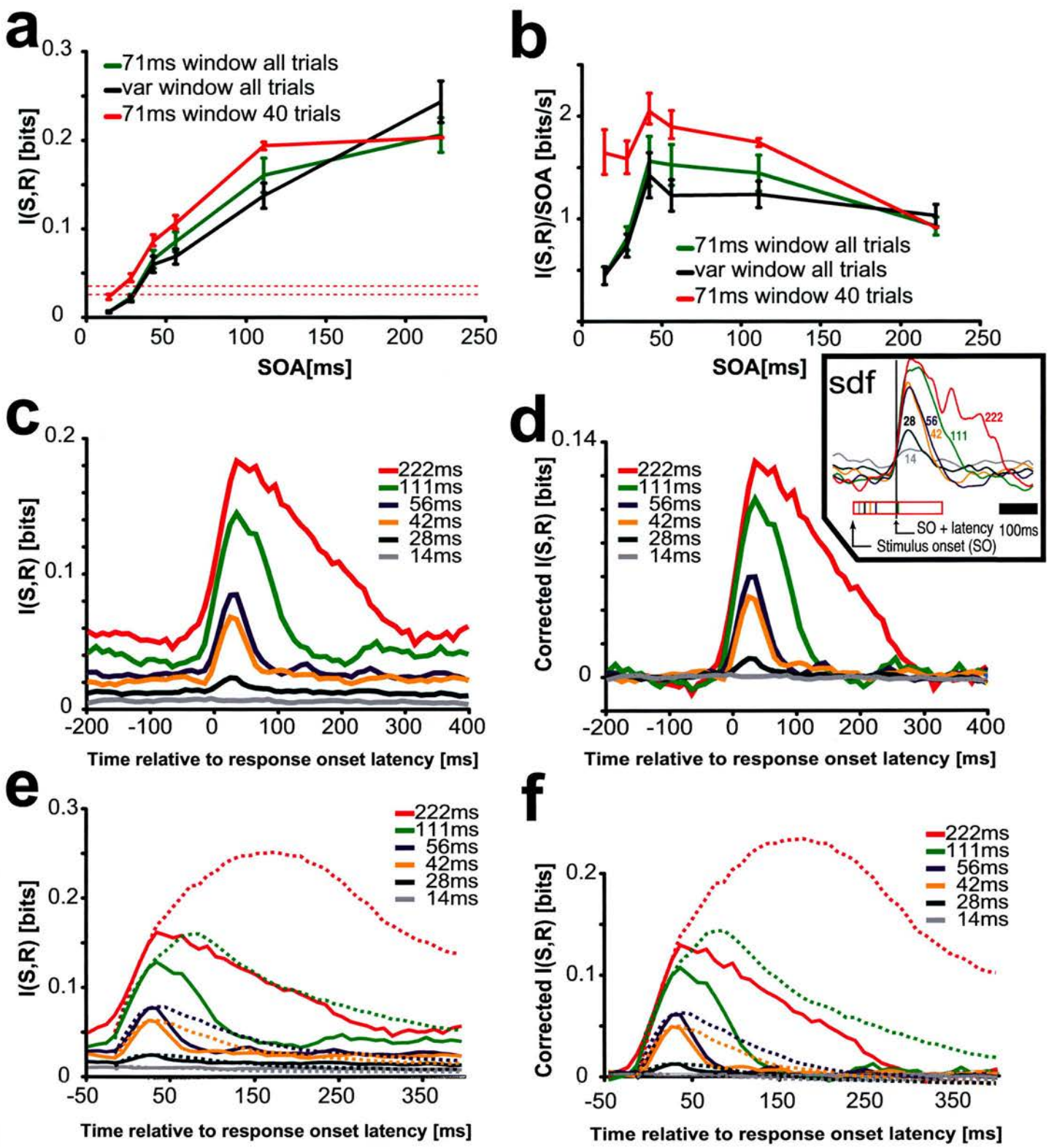


Figure 8.2: I(S,R) as a function of SOA. (a) Green and black curves: Average I(S,R) (±sem) measured based on all trials and the spike-count starting at each cells response onset latency and lasting for 71ms (green) or the variable duration (black, see Table 7.2 col. 4). The values are averaged over all available cells. Red curve: average I(S,R) contained in the 71ms window if only a maximum of 40 trials is considered at each SOA. The error bars for the red line indicate the maximum and minimum average information obtained after averaging the I(S,R) of all cells. The red dotted lines indicate the maximum and minimum average shuffled I(S,R) for 40 trials per stimulus at 14ms/image (see text for details). (b) Same conventions as (a), but the I(S,R) is divided by the SOA. (c) I(S,R) contained in a 50ms spike-count window, which has its centre as indicated on the x-axis, as a function of SOA if all trials are considered. (d) Same as (c) after, for each SOA separately, the average of the I(S,R) contained in the windows centred from -225 to -75 is subtracted to correct for the limited sampling bias. Inset: latency aligned population sdf as in Fig. 7.5. (e) Solid lines as in (c), dotted lines: I(S,R) contained in a window of variable duration, starting at 0ms relative to response onset latency and ending at time x+25ms, where x is the value on the x-axis. The x+25ms adjustment is made so that the I(S,R) contained in the spike count window 0 to 50ms is plotted at the same x-coordinates for both the dotted and the solid lines. (f) same as (e) but corrected as in (d). The same colour code is used from (c) to (f).

cell, at each presentation rate, for each stimulus. The information calculated based on the 40 trial random sub-sample is then averaged over all available cells for a given SOA. This procedure was repeated 30 times, selecting 40 trials at random each time. In all conditions, except for the 111ms/image and 222ms/image, more than 40 trials were available in each conditions (see Fig. 6.4). In condition 111ms/image, 7/34 cells and at 222ms/image, 33/34 cells had less than 40 trials per stimulus. For those cells, the maximum trial number available was used to assess *I(S,R)* in all cases, and the exact same trials were taken for all 30 repetitions of the calculations. For those presentation rates, the range of average *I(S,R)* is reduced compared to what it would have been, if more trials had been available.

The results are shown in Figure 8.2a (red line). Unlike in most other cases, for the red line of Figure 8.2a, error bars do not represent the sem (standard error of the mean), but show the minimum and maximum *average* (averaged over all cells available at each rate) *I(S,R)* obtained during the 30 repeated calculations using 40 trials picked at random each time. As can be seen, the average *I(S,R)* is a stable measure of information that varies little depending on which 40 trials are considered. Also, even if the same trial numbers are taken at all SOA, the information about the stimulus contained in the spike count of the first 71ms of the response decreases with decreasing SOA. Given that the range of means never overlap between presentation rates, no statistics are required to show that all SOA differ with respect to the *average I(S,R)* contained in the first 71ms of the responses. Also, the red curve is approximately parallel to the green curve, indicating that the limited sampling bias of using 40 trials instead of all available trials accounts for ~0.02 bits of information in our data. The only exception is the 222ms/image condition, where trial numbers are equal in both curves, and hence no difference is to be expected.

While the decrease in *I(S,R)* when SOA is reduced can be clearly seen in the values of *I(S,R)*, *I(S,R)* calculated from a limited sample is unsuitable to decide if information is truly non-zero at the fastest presentation rates. This is due to the limited sampling bias. As can be seen in Figure 8.2a, at rapid presentation rates where many trials are collected, calculating information using 40 trials yields to average *I(S,R)* values that are always larger than when *I(S,R)* is calculated using all trials: the range of the red line does not overlap with the values of the green and black line. There is thus no way to know, if the average *I(S,R)* value of 0.0063bits at 14ms/image measured using all trials is different from 0 because just like the value using 40 trials

is higher than that obtained using all available trials, the value calculated using all available trials may still be larger than the value obtained if an infinity of trials had been collected, and the true $p(s,r)$ had been measured. $I(S,R)$ may thus be different from 0 due to true information contained in the cellular responses or to a limited sampling bias alone.

In Chapter 7, classical statistics indicate that for a majority of single cells, and for the population, an ANOVA indicates a difference in mean responses between the 8 stimuli at 14ms/image. This significant difference indicates that information has to be truly non-zero.

An attempt to test if the information value obtained for 14ms/image can be due to limited sampling alone was performed by random trial shuffling. For each cell, 40 trials were taken again at random for each stimulus in the 14ms/image condition, and for each trial, spike counts were taken in the 71ms following the response onset latency of each cell. Responses were shuffled between stimuli, i.e. the 40 trials x 8 stimuli=320 spike counts were randomly assigned to the 8 stimuli. Then $I(S,R)$ was then calculated for each cell between those shuffled responses and the stimuli, and the mutual informations were then averaged over all available cells. The procedure was repeated 30 times. These 'shuffled' $I(S,R)$ values then define a range of $I(S,R)$ values that cannot be trusted to reflect more than a limited sampling bias alone. The two horizontal red dotted lines in Figure 8.2a indicate the range (minimum and maximum) of the average $I(S,R)$ values obtained using this shuffling technique for 40 trials, and should be compared against the red information curve obtained using the same number of trials. All presentation conditions except the 14ms/image condition lie above this 'shuffle range'. The 14ms/image condition lies within and even below the 'shuffle range' and hence, based on $I(S,R)$ alone, it would be impossible to determine if true information about the stimulus was contained in the spike count of the first 71ms of the response of single STSa neurones at that presentation rate. The fact that in the 14ms/image condition the $I(S,R)$ is lower for unshuffled compared to shuffled data is probably related to the fact that the spike count histogram for shuffled data are poorly described by a single Gaussian distribution. The choice of the best fitting Gaussian will thus be less uniquely constraind by the shuffled data compared to the unshuffled data. This will result in less overlap between the Gaussians for different stimuli in the shuffled condition, and thus in higher $I(S,R)$ values calculated using the Gaussian in the shuffled conditions.

Fortunatly, the classical statistics of Chapter 7 indicate clearly, that Information in the 14ms/image conditions is unlikely ($p<0.05$) to be truly zero. 'Classical' statistics are thus superior to information theory when it comes to deciding if the information about a stimulus contained in a response is null or no. Hence for any further measures of *I(S,R)* in this thesis, no attempt will be made at investigating if the *I(S,R)* is above zero or not based on *I(S,R)* itself.

8.3.2. *I(S,R)* per unit of stimulus time

Figure 8.2b illustrates the *I(S,R)* per unit of time. These values were obtained by dividing *I(S,R)* by the SOA and represent the information that can be transmitted per second through the spike count of the average single STSa neurone.

As in Figure 8.2a a limited sampling bias can be clear observed by comparing the *I(S,R)*/SOA considering 40 trials (red curve) with that taking all trials into account (green curve). The difference increases with decreasing SOA, as does the difference in the number of trials considered.

The *I(S,R)* for a maximum of 40 trials (red curve) is relatively flat for SOA between 14 and 111ms/image, but clearly decreases at longer SOA (222ms/image). A 6 SOA repeated measurement ANOVA run on the 30 average mutual information obtained picking 40 random trials each times indicates a significant main effect of SOA ($F(5,145)=808$, $p<0.0001$). A Newman Keuls post-hoc reveals that all SOA yield different average *I(S,R)*/SOA, with the 42ms/image condition yielding the largest average *I(S,R)* by unit of time: 2.05bits/second. The 42ms/image condition also yields the largest *I(S,R)*/SOA if all trials are considered, or if the spikes are counted for the entire response window.

8.3.3. *I(S,R)* contained in a floating window

To analyse the time course of *I(S,R)* relative to the response onset latency, as done by Sugase et al. (1999), *I(S,R)* was calculated for each cell based on 50ms spike count windows taken every 10ms relative to each cell's response onset latency (all trials being considered). The *I(S,R)* was then averaged over all cells, and plotted in Figure 8.2c as a function of the centre of the 50ms window. The *I(S,R)* in windows ending before the response onset latency have values larger than zero, and the conditions with the least trials have the highest pre-response information. This indicates the effect of a limited sampling bias. To correct for this effect, for each SOA

the average *I(S,R)* contained in windows centred between -225ms and –75 were averaged, and subtracted from all *I(S,R)* values for this SOA. The result of that correction is plotted in Figure 8.2d.

The inflection point of the *I(S,R)* curves occurs around t=0, when spikes occuring at and after the response onset latency start to be included in the spike count. The inflection point occurred earlier for conditions with larger peak *I(S,R)* an effect that is typical for large spike count windows: larger responses will have a noticeable effect earlier than small responses, an effect that would disappear for infinitesimally small windows.

The peak of the 50ms *I(S,R)* occurred early in the response, in the 10-60ms window for 222ms, 111ms and 56ms SOA, in the 0-50ms window for 42ms and 28ms SOA and in the –30 to 20ms window in the 14ms SOA. This indicates that most information in conveyed in the beginning of the response, even if the next stimulus occurs much later.

As reported with respect to Fig. 8.2a, the peak *I(S,R)* decreases with decreasing SOA. Also the duration for which the *I(S,R)* appears to be above baseline is dependent on SOA. It is unnecessary to determine the duration for which information remains above baseline, since the interval during which the mean firing rate to different stimuli is different has been previously determined in Chapter 7. Chapter 7 thus indirectly indicated that significant information is available from response onset latency until ~60ms after response onset latency plus SOA.

At 14ms/image, no clear difference is visible between the *I(S,R)* before and during the response of the cell. This is probably due to the true information in the 14ms/image condition being small compared to the limited sampling bias.

To investigate how much information is contained in the beginning of the response, the information contained in the variable spike count window (i.e. during the entire duration of the response) was compared with that contained in 3 shorter windows taken towards the beginning of the response: the best of all 50ms window (i.e. the window at a given SOA with the highest average *I(S,R)* of all 50ms windows as plotted in Fig. 8.2c), the first 50ms window (0-50ms relative to response onset latency) and the first 71ms window. I(S,R) in the briefer windows was then expressed as a percentage of the *I(S,R)* contained in the variable window (see Table 8.1). To test the significance of the differences between the variable and the shorter windows, for each of these windows, the *I(S,R)* obtained for each cell in the brief window was

compared with the *I(S,R)* in the variable window using a 2 windows x 6 SOA, repeated measurement ANOVA. For all brief windows, there was a significant main effect of SOA (all $F(5,70)>39$, all $p<0.001$), never a significant main effect of Window (all $F(1,13)<1.2$, all $p>0.3$, see penultimate row of Table 8.1) and always a significant window x SOA interaction (all $F(5,70)>5$, all $p<0.001$, see the bottom row of Table 8.1). Newman Keuls post-hoc analyses of the SOA x window interaction indicated that at 222ms/image, the variable window always yielded significantly larger *I(S,R)* than the shorter windows, but that at shorter SOA, there was a trend towards the opposite direction, with briefer windows tending to contain more information, but this effect was significant only for the 111ms/image condition comparing the 71ms window against the variable window ($p<0.01$), all other differences were not significant ($p>0.05$). Thus, for the conditions tested in this Chapter, over 69% of information is contained in a 50ms window taken around the beginning of the response.

SOA	*I(S,R)* for the entire response duration	% of the *I(S,R)* contained in the entire response that is contained in:		
		best 50ms	first 50ms	first 71ms
222ms	0.243bit	75%**	69%**	85%***
111ms	0.137bit	105%	100%	117%**
56ms	0.069bit	123%	123%	124%
42ms	0.060bit	114%	114%	110%
28ms	0.021bit	113%	113%	109%
14ms	0.006bit	128%	106%	100%
Main effect of Window		F(1,13)=.01 p>0.9	F(1,13)=1.1 p>0.3	F(1,13)=.22 p>0.6
Window x SOA interaction		F(5,70)=6 p<0.001	F(5,70)=9 P<0.0001	F(5,70)=5 p<0.001

Table 8.1: Distribution of *I(S,R)* over the response duration. The leftmost column indicates the SOA, at its right, the average *I(S,R)* for the entire response window (i.e. the variable response window) considering all trials in indicated. The 3 rightmost columns indicate the percentage of the former *I(S,R)* contained in the spike counts of the windows indicated above. *I(S,R)* in the brief windows were compared with *I(S,R)* in the entire response using using ANOVA and Newman Keuls post-hoc analyses as

indicated in the text. **: Newman Keuls $p<0.001$, ***: Newman Keuls $p<0.0001$. All other Newman Keuls $p>0.05$.

8.3.4. *I(S,R)* as a function of window size

To test how *I(S,R)* changes when spikes are counted over increasingly long periods of time, spikes were counted starting at response onset latency for a duration varying from 10ms to 400ms. *I(S,R)* was calculated and averaged over cells for each spike count window duration. The results are shown as dotted lines in Figure 8.2e as a function of SOA and spike count window size. To allow direct comparison between these *I(S,R)* and the *I(S,R)* contained in 50ms windows both I(S,R)s are plotted against the end of the window minus 25ms. For the 50ms windows, the end of the window minus 25ms is the centre of the window while for the windows starting at 0 and ending at t_{end}, the convention may seem odd, but is necessary to ensure that the *I(S,R)* contained in the window 0 to 50ms – a window used in both methods – is plotted against the same x-axes value (t=25ms).

Different trial numbers are available at different SOA. As in section 8.3.3 the *I(S,R)* were corrected for this limited sampling bias by subtracting the average of the *I(S,R)* contained in 50ms windows centred between -225ms and -75ms. The results are shown in Figure 8.2f.

Fig 8.2e,f illustrates, how the initial rising portion of the *I(S,R)* at each SOA strongly coincides between the two methods (50ms windows and expanding window). This illustrates how the spike counts prior to the response onset latency convey no noticeable information, since for instance a window lasting from 0-30ms contains as much information as a window from –20-30ms.

Table 8.2. illustrates how the information contained in a 50ms window close to the beginning of the response compares to the window starting at 0ms having the best duration, i.e. the duration containing the most *I(S,R)*. Two effects are apparent.

First, it appears, that the window size yielding the largest I(S,R) is approximately identical to the SOA. The fact that counting spikes for a duration longer than the stimulus duration reduces the *I(S,R)* does not indicate, that spikes occurring after response onset latency + SOA do not carry information about the stimulus: indeed, Chapter 7 indicates clearly, that responses carry information for ~60ms longer than stimulus duration (as indicated by significantly different mean firing rates). This fact simply indicates that these later spikes are less informative then

the earlier spikes, and thus dilute the total information calculated based on the spike count.

Second, it appears, that even for relatively long stimuli, a surprisingly large proportion of the maximum I(S,R) is contained in a 50ms window close to the beginning of the response (~60% even for a 222ms stimulus). Hence ~60% of the information is contained in a spike count taken over ¼ of the stimulus duration. Again, this fact does not necessarily indicate that later spikes only contribute 40% of the information, but rather that if a single spike count is taken over longer period of times, only 40% information are added.

To investigate how information accumulates as spike counts are observed for increasingly long periods of the response for a relatively unmasked stimulus, the I(S,R) obtained for spike count windows starting at response onset latency and lasting for a variable duration was calculated as a proportion of I(S,R) obtained in the best window (0-200ms) at that SOA. The result is illustrated in Figure 8.3. Although I(S,R) does not approach its maximum before 200ms after response onset latency, the first 100ms carry 80% of the information and the first 130ms carry 90% of the information.

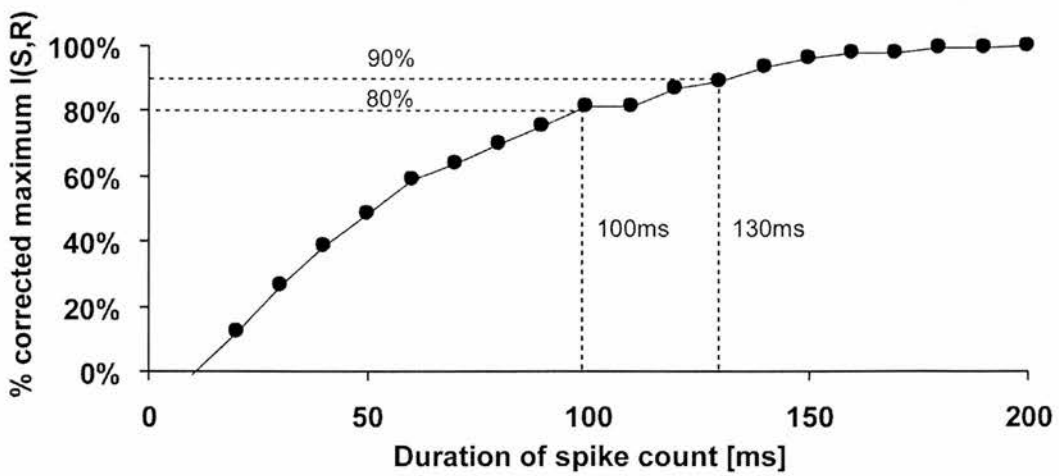


Figure 8.3: Corrected I(S,R) as a function of the end of the spike count window. All spike count windows start at 0ms, and all trials are considered.

		SOA					
		222ms	111ms	56ms	42ms	28ms	14ms
(i) best 50ms window	uncorrected	0.183b	0.145b	0.085b	0.068b	0.023b	0.008b
	corrected	0.128b	0.106b	0.060b	0.048b	0.011b	0.002b
	window	10-60ms	10-60ms	10-60ms	0-50ms	0-50ms	-30-20ms
(ii) best expanding window	uncorrected	0.288b	0.181b	0.087b	0.069b	0.023b	0.008b
	corrected	0.233b	0.143b	0.062b	0.048b	0.011b	0.002b
	window	0-200ms	0-110ms	0-60ms	0-60ms	0-50ms	0-30ms
Percentage (i/ii)	uncorrected	64%	80%	98%	99%	100%	99%
	corrected	55%	74%	97%	99%	100%	97%

Table 8.2: Comparing *I(S,R)* in the best expanding window (as a measure of total information in the response) compared with the best 50ms window. For each SOA the *I(S,R)* contained in the best[8] 50ms window is given both as the uncorrected value, and as the corrected value (corresponding to the peak of Fig. 8.2c and d respectively) together with the corresponding window, which timing is given relative to response onset latency. The same information is given for the best expanding window. Finally, the two values are divided, and expressed as a percentage.

8.4. Conclusions

The single cell recordings of Chapter 7 were analysed according to information theoretical methods to establish how much information regarding the stimulus is present in spike counts of single STSa neurone responses. The aim was also to analyse the time course of information transfer during RSVP.

Results indicate that the mutual information between spike count and stimulus is reduced if SOA is reduced. This is true for all spike count windows used: (1) the entire response window determined in Chapter 7, (2) the window starting at the response onset latency and having the best[9] duration, (3) the first 71ms of the response, (4) the first 50ms of the response and (5) the best 50ms window. This reduction of information when the SOA is reduced has also been described by for visual pattern masking by Rolls et al. (1999). In their study, Rolls et al. measured approximately similar information values for long SOAs: at an SOA of 100ms Rolls et al. measured 0.22 bit of information, and 0.14 bits (corrected value) were measured in the present study for an SOA of 111ms. The reduction of information as SOA is

[8] 'Best window': the window with the largest *I(S,R)* value.
[9] 'Best duration' is the duration for which *I(S,R)* is maximized.

reduced was more pronounced in the present study: at SOA=28ms, 0.011bit of information were measured in the present study (corrected value), corresponding to less than 14% of the information present at 111ms/image; while Rolls et al. information curve reads[10] 0.1bit of information at SOA=28ms, corresponding to ~45% of the information present at SOA=100ms. The more pronounced reduction of *I(S,R)* in the present study supports the idea that RSVP is more challenging for the brain than backward masking at the same SOA. This explains why Uttal (1969a,b) observed how combining forward and backward masking disrupts recognition performance more than expected by forward or backward masking alone.

The duration for which responses contained information about the stimuli was reduced by reducing the SOA, confirming the idea that the following stimulus in an RSVP sequence interrupts the neural representation of a stimulus. The maximum information was contained in a window starting at response onset latency and lasting for a duration approximately equal to the SOA. Including the 60ms of neural persistence described in Chapter 7 does not increase the mutual information. This should not be taken as evidence, that the period of neural persistence does not carry information about the stimulus. Indeed, as demonstrated in Chapter 7, responses carry information during this period, as demonstrated by a significant difference in the mean sdf between the different stimuli.

Why then, does information decrease, if more spikes, which do carry information about the stimulus, are included in the analysis? The problem lies in the method of information measurement: when taking a single spike count over a large window, less informative spikes at the end of the response will *decrease* the overall information contained in the response rather than adding their information towards the total information. Hence the information of the spike count, and not the information of the spikes themselves is measured. To test the contribution of the period of persistence to the information regarding the stimulus, not a single, but at least two independent spike counts need to be used to describe each response: the spike count during the stimulus duration (shifted by the response onset latency) and the spike count in the following 60ms. Knowing the spike count in those two windows separately may be much more informative than knowing only the sum of those two spike counts, or the spike count during the stimulus duration alone. This analysis was

[10] Rolls et al. did not test an SOA of 28ms. The value 0.1bit is obtained by linear interpolation.

not performed in the present chapter because no way has been found to correct for the limited sampling issue: if each response is characterised by 2 spike counts, each spike count combination would be observed less often than if responses are characterised by a single spike count. This would result in different limited sampling biases for both methods – a problem that would prevent a direct comparison of the values obtained. Until this problem is solved, it cannot be determined how much information about the stimulus the ~60ms period of neural persistence can add.

As in other investigations (Tovee et al., 1993; Tovee and Rolls 1995; Heller et al., 1995), a substantial (~60%) proportion of the maximum information contained in the 222ms/image condition was contained in the first 50ms of firing. An even larger proportion was contained in the first 50ms at 111ms/image (~80%). For SOA of 56ms and shorter, essentially all the information is contained in the first 50ms of the response.

For a SOA=222ms, information continued to rise for up to 200ms after stimulus onset, but the increase of information started to reach a relative flat plateau much earlier, with 80% of the information being available after 100ms of response, and 90% being available after 130ms. This finding is similar to the findings of Gershon et al. (1998) and Tovee et al. (1993), both of which show that even for longer stimuli 80% of the information is accumulated at 200ms after stimulus onset, which given an average latency of ~100ms in IT would correspond to 100ms post latency in the present experiment.

The Information per unit of time had its maximum of ~2bits/second at 42ms/image, and not at longer stimulus durations. This suggests, that for a neurophysiologist, the best testing rate for cell in STSa is around 42ms/image: this is where the most can be learned about the stimulus preferences of a neurone per unit of stimulus time. In addition, surprisingly, it suggests that the most bits of information about the outside world could be extracted in a limited amount of time, if a new image occurred every 42ms. This would suggest, that saccades should occur at approximately that rate. Yet saccades usually occur about once every 200-300ms, at a rate at which much smaller information rates were measured in the present experiment (~1bit/s). It may be that rather than optimising information transmission rates, the system evolved to extract as much information as possible from one saccade before moving to the next saccade. Also it may be that under natural viewing conditions, were the images in successive saccades are statistically dependent, less information

may be gained by increasing the saccade frequency. Our findings in the 222ms/image condition and those of Gershon et al. (1998) and Rolls et al. (1999) suggest that about 100-200ms are sufficient to reach the plateau of the information regarding a stimulus. This explanation would better explain the observed inter-saccade interval of typically 300ms

9. Experiment 2: A psychophysical investigation of memory and detection performance in RSVP sequences without inter-stimulus gaps

9.1. Introduction

The single cell investigation of experiment 1 revealed that single cells in the macaque STSa display significant stimulus discrimination for images presented in a continuous RSVP sequence at 14ms/image. Single cell stimulus discrimination at such rapid presentation rates raises the question of what the monkeys perceived during these RSVP sequences: could they perceive the stimuli their single cells responded to?

To answer this question, in retrospect, it would have been best to train monkeys to report the perception of a certain target in the sequence while electrophysiological recordings were being performed[11]. Yet, it is known that humans and rhesus monkeys have very similar visual systems. Both species respond in similar ways to binocular rivalry (Leopold and Logothetis, 1996; Sheinberg and Logothetis, 1997; Logothetis, 1998; see section 2.2.4). With respect to the speed of visual processing in the two species Fabre-Thorpe et al. (1998a,b) could show that monkeys and humans take comparable times to decide if a visually presented scene contains an animal. Most importantly, in masking experiments, both species respond similarly to metacontrast masking (Bridgeman, 1980) and their performance at detecting a particular target when it is followed by a pattern mask is virtually identical (Kovacz et al., 1995, results illustrated in our Fig. 3.17). Indeed, authors have used monkeys for electrophysiological investigations of the visual system under masking conditions, and directly compared these results with psychophysical measures of human

[11] This is clearly a worthwhile - although lengthy - experiment to pursue in the future. It would allow us to compare single cell performance and perceptual report on a trial by trial basis, in a way similar to the excellent series of experiments of Shalden, Newsome, Celebrini, Britten and Movshon, which gave profound insights into the perception of motion. (Shalden and Newsome, 1996; Celebrini and Newsome, 1995; Shalden et al., 1996; Britten et al., 1996).

perception under similar conditions (e.g. Rolls et al., 1994a, 1994b, 1999; Macknik and Livingstone, 1998). Accordingly, we opted to draw on human observers to measure detection and memory performance in RSVP sequences using the exact same stimuli presented in the electrophysiological investigations of Experiment 1 (see Fig. 9.1)

The existing literature on human performance in RSVP sequences using naturalistic images (i.e. images that resemble those found under natural conditions, including computer generated images and cartoons and line drawings) is unanimous about the fact that both perception and memory for images decreases with decreasing SOA, but gives often contradictory information about the point at which detection or memory performance reaches chance level: some authors find for instance significant memory at 125ms/image (Potter, 1976 and Potter and Levy, 1969) while others report no explicit or implicit memory for stimuli at similar presentation rates (Subramaniam et al., 2000). Subramaniam et al. (2000), attribute such differences mainly to the nature of the stimuli (photographs were used by Potter, 1976 and Potter and Levy, 1969; and line drawings by Subramaniam et al., 2000). Considering these contradictions in the literature, it was imperative to measure human psychophysical performance using the same stimuli and conditions used in Experiment 1.

Figure 9.1 illustrates how the very same stimuli used in Experiment 1 were used in Experiment 2. In Experiment 1 (Figure 9.1b), 23 single cells had been recorded at 14ms/image. For each single neurone, a specific test set of 8 stimuli had been used: the 'best' stimulus and the 7 remaining 'rest' stimuli. In the electrophysiological recordings, the monkey had only to fixate the RSVP sequence, and thus, no systematic selective attention for one of the stimuli was generated. The goal of the separate psychophysical experiments presented in this chapter (Fig. 9.1a, c) was to define a bracket of performance indicating how much perception of individual images occurred in the physiological experiments. The performance in a detection (Fig. 9.1a) and a memory (Fig. 9.1c) task was used to estimate the upper and lower limit of this perception respectively. Some stimuli may be easier to detect than other, and therefore the performance for all 23 stimulus sets containing 8 stimuli used previously in (b) was measured, and for each stimulus set the stimulus that had been

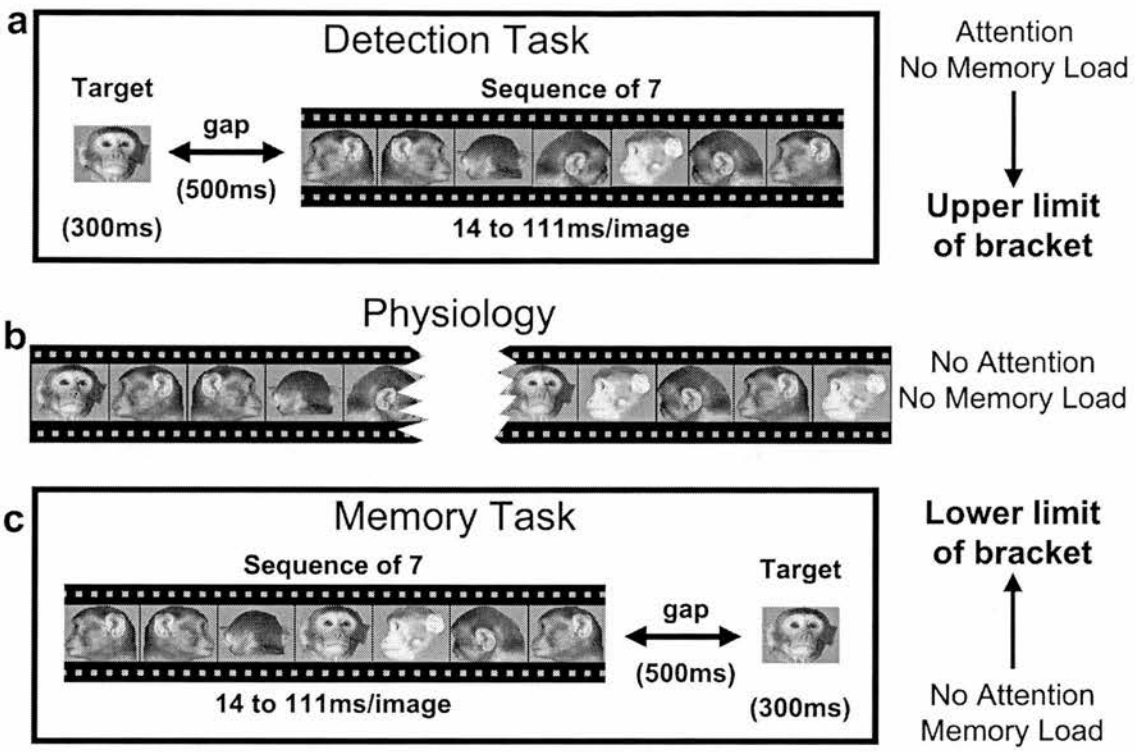


Figure 9.1: Illustration of the experimental design of Experiment 2 (panel a and c) as compared to Experiment 1 (panel b). The movie strips represent RSVP sequences, the white gap in the strip in (b) represents the fact that up to 720 stimuli were presented. See text for details.

best for a cell was chosen as the 'target', and the other 7 stimuli used for that cell as 'distracters'. In the detection task (Figure 9.1a), each target was presented for 500ms, followed after a 500ms gap by an RSVP sequence of 7 images either containing only the 7 distracters, or containing the target in position 3-5 surrounded by 6 distracters. The human subjects signalled perception of the target in the sequence by pressing a key. Trials with each of the 23 stimulus sets were intermixed and presentation rate (111, 56, 42, 28, 14ms/image) for the test sequences varied randomly between trials. In this detection task, subjects selectively attended to a target thereby probably privileging the processing of the target at the expense of other stimuli (Duncan, 1998; Desimone, 1998; Treue and Martinez Trujillo, 1999; Chun and Potter, 1995). This selective attention, present in the detection task but absent from the physiological test situation, makes the behavioural performance in the detection task an estimate of the upper limit of the perception possible in the RSVP sequences used in the physiological testing situation (Fig. 9.1b). Another task was therefore needed to give a *lower* limit of the perception possible in the physiological recordings. In the memory task (Fig. 9.1c), the RSVP sequence was shown first, followed by a 500ms gap and a

300ms target. The subjects signalled their memory of having perceived that target in the preceding sequence by pressing a key. In this task, the subjects could not selectively attend to the target during the RSVP sequence, since the target was still unknown to them at that point. In addition, in this task, the stimuli in the sequence had not only to be perceived, but also remembered for a brief interval. This additional memory requirement in the absence of selective attention makes the behavioural performance in the memory task an estimate of the lower limit of the perception possible in the physiological testing situation.

The goal of this experiment was to investigate if the significant neuronal stimulus discrimination observed in the STSa of macaque monkeys in 14ms/image RSVP sequences would enable a subject to perceive the stimuli. Perception will be used in a behavioural and operational definition: a stimulus is perceived if the subject performed above chance in a psychophysical task based on the stimulus. As will be discussed in 9.4.4., above chance performance in a psychophysical task does not necessarily mean that the perception was conscious: there are many cases in the literature in which above chance performance can occur while subject report no conscious awareness of the identity of the stimulus (Weiskrantz, 1996; Stoerig & Cowey, 1997; Holender, 1986; Merikle, 1992; Dehaene et al., 1998). Perception hence is to be differentiated from conscious perception.

9.2. Methods

9.2.1. Subjects, materials and procedure.

Five human subjects with normal or corrected to normal vision were used for this experiment. Each subject was seated in front of the same apparatus used for the electrophysiological experiments (see Figure 6.1) with the exception that eye movements were not measured. The same level of room lighting was used, and subjects were seated at the same distance from the screen (~50cm) as in Experiment 1. Figure 9.1 illustrates the procedures used in the psychophysical experiments, and how they compare to the procedures in the electrophysiology. In 50% of the trials the RSVP sequences contained only the 7 distracters in random order; in 50% of the trials the sequences contained 6 distracters picked at random from the possible 7 and the target in position 3, 4 or 5 of the sequence. The memory and the detection task were performed on different sessions; 3 subjects performed the detection task first and the

memory task second, for 2 subjects it was the other way round. The 23 stimulus sets were tested 16 times each (8 target present, 8 target absent trials) in each task at presentation rates of 14, 28, 42, 56 and 111ms, totalling to n=16x23x5x2=3680 trials per subject. Presentation rate and stimulus were interleaved randomly on a trial by trial basis. The data from the 5 subjects were pooled. Subjects included the author, two other investigators involved in the research (DX, PF) and two naïve subjects (TJ and RE).

9.2.2. Data analysis

The trials from all 5 experimental subjects were pooled for each stimulus, yielding a signal detection table (McNicol, 1972) for each stimulus (See Table 9.1).

	Response	No Response
Target present	**13** **Hit**	**27** **Miss**
Target absent	**4** **False alarm**	**36** **Correct rejection**

Table 9.1: Signal detection table, with values taken from the experiment at 14ms/image in the detection task. A total of 80 trials (5 subjects with 16 trials each) for each stimulus is composed of 40 target present, and 40 target absent trials. When the target was present in the sequence, in this example, the subjects pressed the button (= 'response') 13 times, indicating that they believed that they had seen the target in the sequence, and refrained from responding (= 'no response') 27 times, indicating that they believed that they did not see the target in the sequence. Those cases are called 'Hit' and 'Miss' respectively. When the target was absent, the subject responded 4 times ('False alarm') and refrained from responding 36 times ('Correct rejections').

The signal detection table was then analysed according to two methods.

9.2.2.1. Proportion correct

Proportion correct (*p(c)*) was calculated for each stimulus separately as described in Equation 9.1 as the number of hits plus the number of correct rejections divided by the total number of trials for the stimulus (i.e. 80 = 5 subjects x (8 target present + 8 target absent trials per subject per stimulus).

$$p(c) = \frac{hits + correct\ rejections}{total\ number\ of\ trials} \quad \text{[Equation 9.1]}$$

This proportion correct score can be statistically tested against the performance expected by chance (p=0.5) by the application of the cumulative binomial distribution, as mentioned in the text.

9.2.2.2. Mutual information

In analogy to the analysis of the single cell recordings in Chapter 8, an information theoretical analysis of the psychophysical data was performed separately for each stimulus, as a direct application of the Shannon Equation (Shannon, 1948, our Equation 9.2).

$$\mathrm{I(S,R)} = \sum_{s}\sum_{r} p(s,r)\log_2 \frac{p(s,r)}{p(s)p(r)} \quad \text{[Equation 9.2]}$$

(In Equation 9.2., *S* is the Stimulus variable which can take the values "present" or "absent", and *R* is the response variable, that can take the values "yes" and "no", and *I(S,R)* is the mutual information between *S* and *R* expressed in bits. $\log_2$ refers to a modified logarithm rule, which is identical to a normal logarithm in base 2 except that $\log_2(0)=0$).

The mutual information between the stimulus (S) and the response (R) indicates how much the presence/absence of the stimulus in the sequence helps predict the response of the subject, and the other way around, how much the response of the subject informs us about the presence/absence of the stimulus in the sequence. This information will vary between 1 bit (perfect 1:1 relation between the two variables) and 0 bit (no relation at all). A mutual information value of 1 would typically be achieved if the subjects always pressed the button if the image was present, and never pressed the button when it was absent from the sequence, or theoretically when the subject never presses the button when the target is present, but always presses it when the target was absent. A mutual information value of 0 on the other hand is achieved when the subject press the button equally often when the target is present or absent from the sequence. This analysis is unusual for psychophysical experiments, but is interesting in comparison with the information theoretical analysis of the single cell data.

9.2.2.3. Reaction time.

In addition, for hit trials, the reaction time between the onset time of the second occurrence of the target (i.e. the target in the sequence for detection task, and the target presented after the sequence in the memory task) to the time of button press was measured, averaged over all subject and trials using the same stimulus, and analysed using parametric ANOVA.

9.3. Results

9.3.1. Proportion correct decisions

The results from the psychophysical experiment were first analysed with respect to the proportion correct responses [*p(c)*] for each subject (see Fig. 9.2a) using an ANOVA (see Table 9.2). All main effects and interactions were significant at $p<0.05$. Of particular interest is the fact that the main effect of task was significant at $p<0.01$ with detection better than memory, and that the main effect of SOA was extremely significant ($p<0.001$). A post-hoc investigation (Newman-Keuls Test, $p<0.05$) indicated that all data pairs in Figure 9.2a are different, except for the pairs d14=m14, d14=m28, d28=m111 (d14 stands for detection task at 14ms/image, m28 for memory task at 28ms/image etc.).

	df Effect	MS Effect	df Error	MS Error	F	p-level
Task	1	3.215	4	0.050	63.66	0.001337
SOA	4	1.724	16	0.010	157.30	1.26E-12
Stimulus	22	0.428	88	0.033	12.75	7.09E-19
Task x SOA	4	0.157	16	0.011	13.84	4.6E-05
Task x Stim	22	0.046	88	0.015	3.03	0.000122
SOA x Stimulus	88	0.024	352	0.009	2.71	4.51E-11
Task x Stimulus x SOA	88	0.010	352	0.007	1.44	0.010518

Table 9.2: Statistical analysis of *p(c)* values in the psychophysical data. A 2 task (detection vs. memory) x 5 SOA (14 to 111ms/image) x 23 Stimuli (1-23) ANOVA with all variables as within subject factors and 5 subjects was used.

There was a significant effect of stimulus[12], indicating that some stimuli were easier than others. Figure 9.3 illustrates two examples of easy and two examples of difficult stimuli as indicated by the *p(c)* in the memory task at 14ms/image. To investigate if a stimulus that was hard at one SOA/task, is also hard at an other SOA/task, *p(c)* was calculated using the pooled data of the 5 subjects for each of the 23 stimuli x 2 task x 5 SOA combination, and the 23 *p(c)* values so obtained at each task/SOA could be correlated pair-wise with the 23 *p(c)* in all other task/SOA. All pair-wise correlations[13] were positive, and 43 out of 45 correlations were significant at $p<0.05$ and only 2 (m14 vs. d111 and m14 vs. m111) were not, indicating that if a stimulus is difficult in one task/SOA, it will be difficult in an other task/SOA.

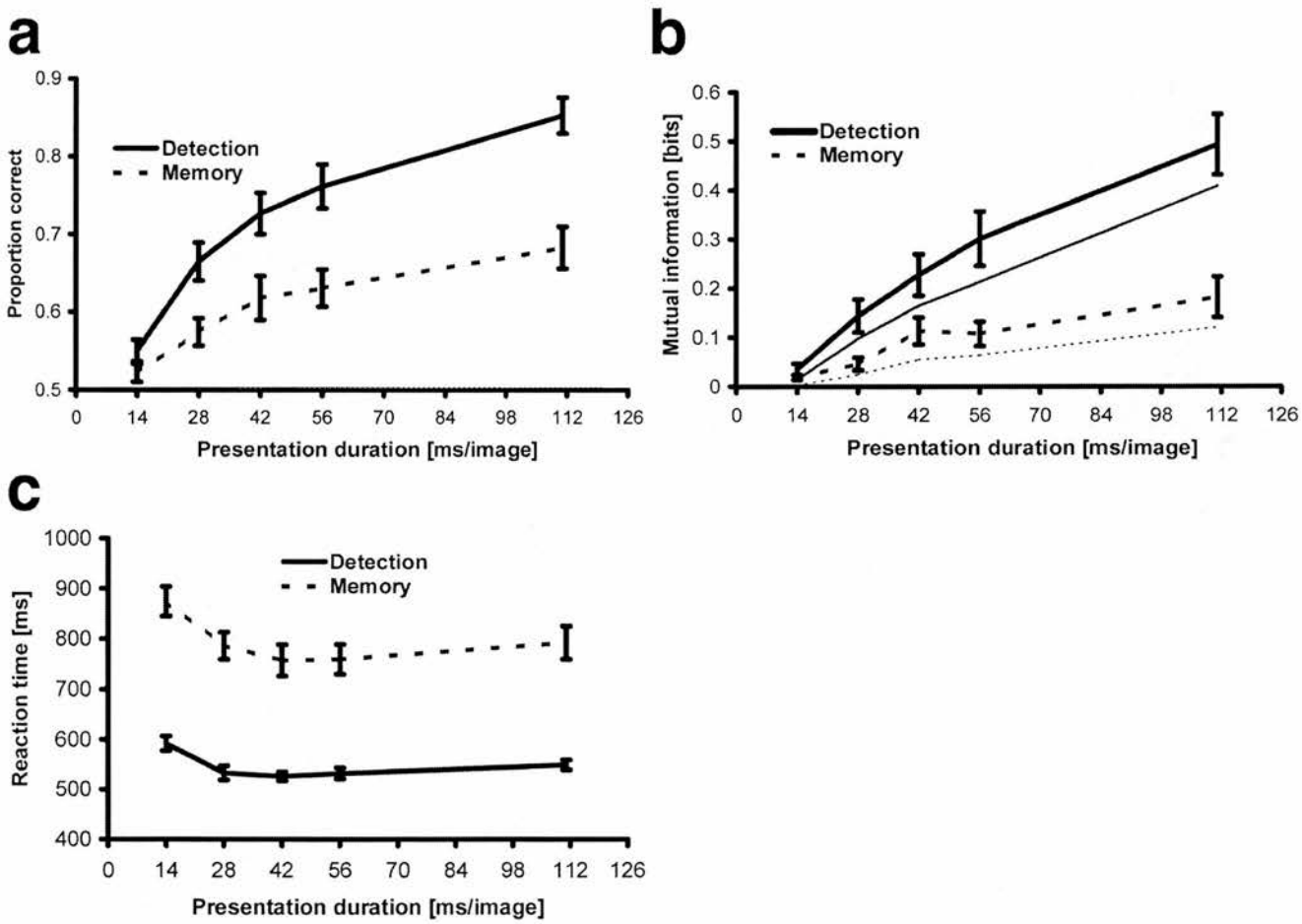


Figure 9.2: Results from the psychophysical investigation. (a) Mean (±s.e.m. of the 23 stimuli) proportion correct (*p(c)*, see methods) as a function of presentation

[12] 'Stimulus' refers to a specific target stimulus together with its 7 distracters identical to the best and rest stimuli, which were chosen in Experiment 1 based on the response of a single cell.

[13] In this thesis, correlation always refers to the Pearson product-moment correlation coefficient, and the correlation *r* is tested using the student-*t* distribution, to indicate if it differs from r=0.

duration and task (memory vs detection). (b) same as (a) but for the mutual information between response and stimulus presence expressed in bits of information. The thick lines represents the average of the information calculated stimulus by stimulus, while the thin lines represent the information calculated on the pooled trials of all stimuli. See text for explanation of the difference between the thick and thin lines. (c) same as (a) but for reaction time.

For all further analysis, the data from all 5 people are pooled for each stimulus, rate and task separately, yielding 80 trials per stimulus. A binomial distribution with 0.5 success probability indicates that for $p<0.05$, at least 47 correct decisions must be reached out of 80 trials. Based on this criterion, even at the fastest presentation rate, 4 stimuli could be memorised, and 6 detected, which is more than expected by chance (see Table 9.3; since in each case, 23 binomial tests were performed for 23 stimulus sets, we tested if 4 significant tests out of 23 performed at $p<0.05$ could be due to chance. The result is negative: a binomial indicates that the probability of 4 or more successes out of 23 repetitions at $p=0.05$ each is less 0.005.) If the responses are analysed irrespectively of the stimulus (i.e. all 23 stimuli are pooled), there were more correct choices at each rate than expected by chance: in the most difficult task, the memory task at 14ms/image, 960 of the 1840 responses were correct (Binomial $p<0.03$), indicating that both memory and detection was above chance even at the fastest presentation rate.

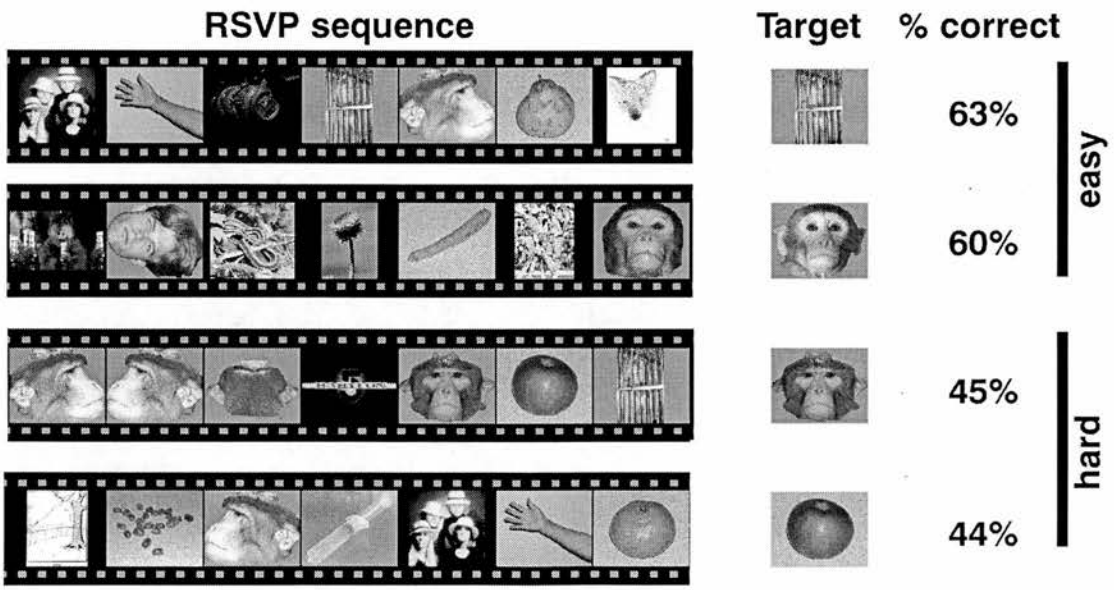


Figure 9.3: Example of easy and hard stimulus sets and the relevant proportion correct responses in the memory task at 14ms/image. Left, an example of each RSVP sequence is shown, that could contain the target (p=0.5) or not (p=0.5), followed by

the target (centre). The subject had to respond whether the target was in the sequence or not. The percentage correct decisions calculated over all 5 subjects is given at the right of the target. The two top examples are examples of stimuli associated with the best, the two lower examples with the worst performance at that presentation rate. Note for the second example from the top the high performance despite the presence of a highly similar distracter in the sequence. (The target was present in half of those sequences, in which case, one of the 7 distracters is taken out of the 7 image sequence).

	Detection	Memory	Detection	Memory
14ms/image	6	4	26%	17%
28ms/image	15	9	65%	39%
42ms/image	19	13	83%	57%
56ms/image	22	13	96%	57%
111ms/image	23	16	100%	70%

Table 9.3: Number and percentage of stimulus sets being in the top 5% of the binomial distribution based on 80 trials and a p=0.5 assumption.

9.3.2. Reaction time.

The reaction time for hit-trials is illustrated in Figure 9.2c, and indicates that the 14ms/image condition resulted in longer RT than the other conditions, and that the memory task resulted in longer RT than the detection task. Hence, the more difficult conditions (i.e. having the lower percent correct performance) are also the ones taking more time to solve. The decrease in proportion correct performance as SOA decreased was thus not due to a speed-accuracy trade off.

For each 5 SOA x 2 task x 23 stimulus combination, the RTs were averaged over all hit trials of all 5 subjects. These averages were then analysed in an 5 SOA x 2 Task (memory vs detection) ANOVA (both factors as within subject variables). The analysis revealed a main effect of Task ($F(1,20)=26180.40$, $p<0.0001$), with the memory task yielding RT ~250ms longer than the detection task, and a main effect of SOA ($F(4,80)=6836$, $p<0.0001$) but no interaction ($F(4,80)=1.14$, $p>0.3$). A Newman-Keuls post-hoc analysis of the effect of SOA indicated that for each task, all SOA except for 14ms/image do not differ significantly ($p>0.05$), indicating that the RT at 14ms/image was longer, but all other RTs were equal.

Correlating the 23 RTs and 23 proportions correct obtained in each of the 10 conditions (2 task x 5 SOA) pair-wise with each other revealed a negative overall relationship, indicating that a difficult stimulus takes longer to identify. This is especially true for the memory task (see Table 9.4).

		RT									
		d14	d28	d42	d56	d111	m14	m28	m42	m56	m111
Prop. Correct	d14	0.01	-0.21	-0.37	-0.33	-0.24	-0.39	**-0.47**	-0.28	**-0.49**	**-0.59**
	d28	0.35	0.02	**-0.45**	**-0.46**	**-0.63**	-0.29	**-0.57**	**-0.48**	**-0.59**	**-0.71**
	d42	0.28	0.16	-0.28	-0.41	**-0.60**	-0.17	**-0.50**	**-0.58**	**-0.56**	**-0.65**
	d56	0.32	0.18	-0.21	-0.37	**-0.63**	-0.08	**-0.49**	**-0.55**	**-0.52**	**-0.67**
	d111	0.24	-0.04	-0.25	-0.24	-0.37	0.11	-0.26	**-0.48**	-0.24	**-0.56**
	m14	-0.01	-0.32	-0.17	-0.34	-0.35	**-0.52**	-0.38	-0.41	-0.25	**-0.62**
	m28	0.08	0.03	-0.33	-0.27	**-0.49**	**-0.45**	**-0.63**	**-0.52**	**-0.67**	**-0.77**
	m42	0.18	0.06	-0.23	-0.17	-0.42	-0.38	**-0.60**	-0.38	**-0.51**	**-0.67**
	m56	0.19	0.23	-0.26	-0.38	**-0.60**	-0.38	**-0.61**	**-0.52**	**-0.60**	**-0.77**
	m111	0.25	0.04	**-0.48**	-0.28	**-0.44**	-0.20	**-0.59**	**-0.48**	**-0.59**	**-0.70**

Table 9.4: Correlation matrix of the reaction time (RT) and the proportion correct in the 10 tasks across the 23 stimulus sets. For each stimulus, RT and prop. correct are averaged over all trials and subject. Correlations that are significant ($p<0.05$) for 23 entries (df=21) are in bold characters. 'm' denotes the memory task, 'd' the detection task, followed by the SOA in ms.

9.3.3. Information

Mutual information was measured for the psychophysical performance based on the signal detection table (see methods, in particular Equation 9.2) for each stimulus pooled over the 5 subjects. Information indicates how well the response of the subject predicts the occurrence of the stimulus and vice versa, how well the occurrence of the stimulus predicts the response of the subject (hence the term '*Mutual* information'). As such, information is strongly related to proportion correct, with the difference that information is positive when the proportion correct deviates from 0.5 in *either* direction: both $p(c)=0$ and $p(c)=1$ are associated with a I(S,R)=1bit (for a yes/no response)– in both cases the response perfectly predicts the presence of the stimulus: at $p(c)=0$, not pressing the button predicts the presence of the target in the sequence, while at $p(c)=1$, pressing the button indicates that presence of the target. The goal of calculating information is to analyse the psychophysical data using the

same methods as the neurophysiological data, to open the way to a comparison that will be made in the next Chapter.

Figure 9.2b illustrates the effect of presentation rate and task on mutual information, while Figure 9.4 illustrates the strong relationship between information and proportion correct. Clearly, information decreases with decreasing SOA, and the memory task was more difficult than the detection task.

Information was above zero at all rates, but this is not a reliable indication of above chance performance. This can be illustrated in a thought experiment. If an inadvertent experimenter forgot to turn on the screen on which stimuli are presented, subjects would now respond randomly. If an infinity of trials was collected, *p(c)* would be equal to 0.5 correct and information would be 0. If on the other hand, only 16 trials are collected for each subject, performance will fluctuate. One subject may have *p(c)*=7/16=0.44, one may have *p(c)*=9/16=0.56 and yet another *p(c)*=8/16=0.5 etc. This will result in an *average* very close to 0.5 for proportion correct, because the error (i.e. deviation of sample mean from population mean) is symmetrically randomly distributed, and averages at 0. For information, both a performance of 0.44 and a performance of 0.56 will yield positive information. Hence if true information is zero, the error in the information will always be positive (since information cannot be negative), and inflate the average. Because of that, information based on the pooled trials for all stimuli was also calculated (see thin lines in Fig. 9.2b). This information is lower than the average of the stimulus by stimulus information, in accordance with the above thought experiment, but even information calculated in this fashion over 1840 trials still remains above zero.

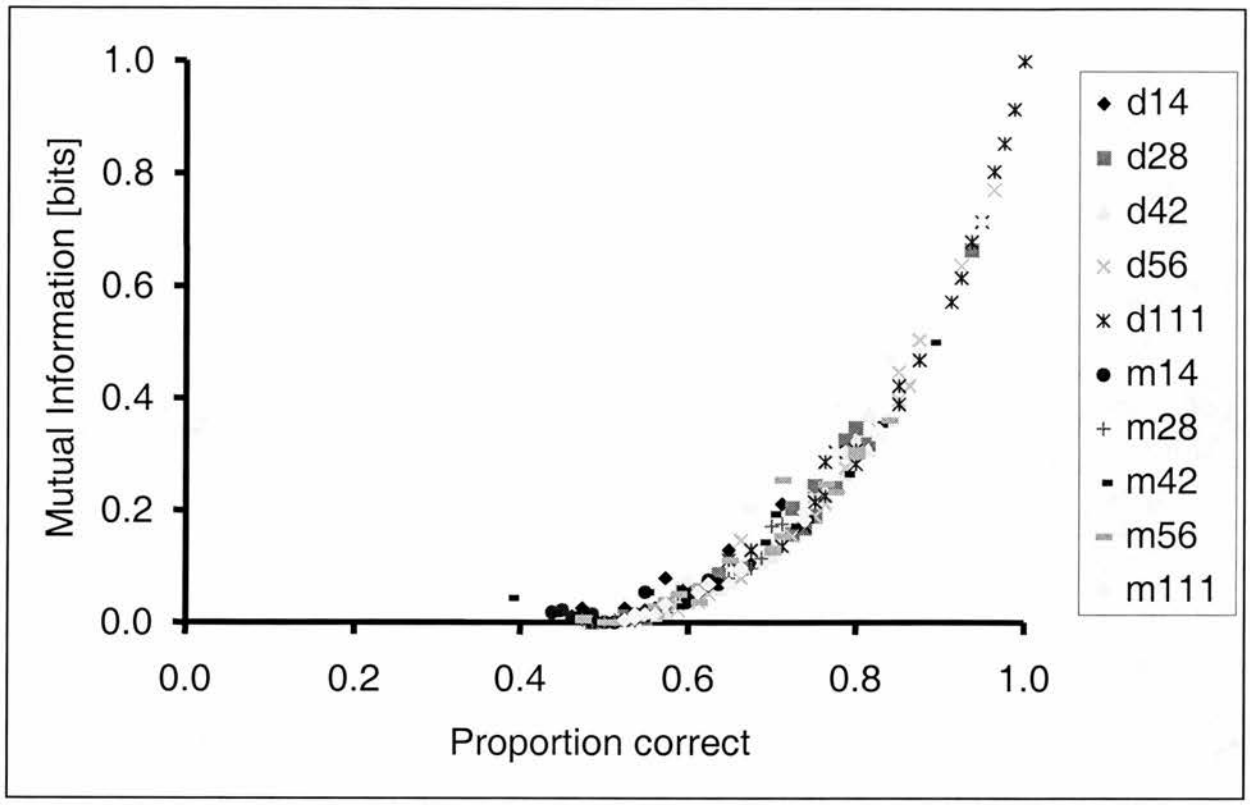


Figure 9.4: Relation between proportion correct and mutual information for each stimulus set for the different task x SOA combinations. The tasks are labled with "d" for detection and "m" for memory followed by the SOA in ms.

9.4. Discussion

The main findings of this experiment is that as RSVP presentation duration is decreased from 111ms/image to 14ms/image, both memory and detection performance is reduced but never abolished: even at 14ms/image both recognition and memory performance is above chance as assessed based on a Binomial test of correct responses.

9.4.1. Why is memory performance worse than the detection performance?

Along with other authors, we found memory performance to be worse than detection performance in RSVP sequences (e.g. Potter and Levy, 1969; Subramaniam et al., 2000). In some cases, a stimulus would thus have been perceived, but less than a second later, the subject has forgotten that perceptual experience. Four lines of explanation of the difference in performance between the detection and the memory task will be discussed.

Visual priming refers to the fact that seeing an image will facilitate its detection in a subsequent sequence. In the detection task, but not in the memory task, the target was presented directly before each sequence. While tempting at first, explaining the difference in performance by the visual priming induced by the visual presentation of the target image just before the sequence is probably incorrect: visual priming is a long lasting effect (e.g. Bar and Biederman, 1998 for priming of image detection under backward masking, and Subramaniam et al. 2000 for priming of image detection under RSVP conditions). If a stimulus is presented, it will prime detection of that stimulus for hours. For the very first trial with each stimulus there may thus be priming only in the detection task, but already by the end of the first trial, in both the detection and the memory task, the target will have been seen once outside of the RSVP sequence. The 15 following trials will thus benefit from comparable amounts of priming.

Subramaniam et al. (2000) has show how long lasting visual priming can be, and hence that it is not important whether the prime and target are presented in immediate succession or not. Subjects had to detect a target stimulus in an RSVP sequence, much like in the experiments presented in this chapter. The authors compared 3 variants of the task. Their RSVP sequences were always composed of 32 line drawings presented at rates of 72ms/image and 126ms/image. In the first task, the target was specified by name ("Ladder") followed by the RSVP sequence that contained the target in 50% of the cases (no visual priming condition). In this variant of the task, subject never saw the target *image* but only the name of the target outside of an RSVP sequence. Seeing the target image as a filler item (i.e. non-target) in a previous RSVP sequence at 72 or 126ms/image did not improve the subjects' performance at detecting the target in a subsequent RSVP sequence. In the second variant of the task, the target was also specified by name, but before the start of the experiment, half of the target images were shown to the subjects for 5s each (mass visual priming condition), without specifying that they would be targets later on. In the third task, most resembling the detection task used in the experiment reported in this chapter, the target was specified before each sequence not by name, but by showing the target image itself for 72 or 126ms/image, not followed by a mask (visual target condition). The authors observed that both the visual target condition and the mass visual priming condition produced better hit rates than the no visual priming condition. The hit rates were 61% hit for the no visual priming condition, 71% for the

visual target condition and 79% for the mass priming condition. Hence, mass priming before the start of the experiment was even more effective than a specific priming before each task. This indicates, that overall (except for the first trial), priming should be equal in the memory and detection task used here and thus cannot account for the difference in performance.

Indeed, our subjects had been exposed to the targets not only in the experiment, but also prior to the experiment, since most of the subjects had been involved in the electrophysiological recordings using the same stimuli. Both memory and detection tasks thus benefit from visual priming, which may explain the relatively high performance in all tasks. Indeed, the same was true for the monkey subjects: over the months of training and recording, they also were exposed repeatedly to the stimuli used in the experiments. In particular, each time a monkey's gaze left the 5° fixation window, the sequence was aborted, leaving the last stimulus unmasked and potentially resulting in strong visual priming. The fact that the monkeys too benefited from visual priming makes the human and monkey data comparable.

A second explanation of the superior performance in the detection task, as mentioned in the introduction, rests on attention, and the **attentional blink** phenomenon. In Chapter 5, Attentional Blink has been described (AB; Raymond, Shapiro and Arnell, 1992; Chun and Potter, 1995), and refers to the fact than in RSVP, when a subject is actively searching for a particular target, finding the target at time t=0 in the sequence will reduce the probability of finding another target in the interval t=200ms-500ms. Interestingly, AB generally does not affect the first stimulus following the target: a second target presented directly after the first (i.e. 100ms after the first) will be processed just as well as the first target. It is only in the interval t=200-500ms that stimuli are 'blinked', as if the subject closed his eyes 200ms after finding the target to protect the processing of the target from the masking effect of subsequent distracters. It appears that the process that protects the stimulus processing against other stimuli may take some time to "close the gate" to the other stimuli, which is why stimuli occurring shortly after the first (e.g. within 100ms) are still processed. Due to the lack of neurophysiological investigations of AB, it is hard to know how AB benefits the cellular responses to the target, and how it exactly compares to other forms of attention that have been measured at the single cell level (e.g. Duncan, 1998; Desimone, 1998; Treue and Martinez Trujillo 1999). Indeed, AB

experiments never measured the benefit of AB for detecting the first occurrence of a target, but only the cost on detecting a second target. While it is tempting to assume that the cost for the subsequent target is a benefit to the first target, there is no definitive proof for that benefit. The problem is that memory and selective attention are inevitably confounded in RSVP experiment. Either the target is specified before or after the sequence. If it is specified before, performance is measured with the potential benefit of attention and without the potential cost of requiring memorisation of the entire RSVP sequence. If the target is specified after the sequence, there is no selective attention, but there is necessarily the cost of memorising the RSVP sequence. Hence, differences are never clearly attributable to the benefits or costs of one or the other. Our experiment is no exception: the detection superiority may be due equally to selective attention or memory or both.

Which leads to **memory decay** as being another likely candidate for explaining the difference between the two tasks. In a detection task, what needs to be memorised is the initial target, which is presented for 300ms followed by a 500ms gap. Neural activity under those circumstances will be strong and of long duration, and possibly followed by delay activity (see Chapter 3). As soon as the target reappears in the RSVP sequence, the behavioural response can be initiated, requiring no memorisation of the occurrence of the target in the RSVP sequence. In the memory task on the other hand, the occurrence of the target in the sequence must be memorised until the target is defined after the end of the sequence, and the occurrence of the target has to be memorised along with the other images in the sequence. In Experiment 1 we demonstrated how the duration of neuronal activity is curtailed to SOA + 60ms under those conditions, and how, especially for SOA<60ms, the amplitude of the response is also reduced. It is almost trivial to state that such reduced responses are more likely to produce a weaker memory trace that will decay more rapidly than that caused by the single, long, unmasked target presentation before the sequence in the detection task.

Subramaniam et al. (2000) speculated that the duration of neural activity may be crucial for the creation of the memory trace necessary to solve the memory task. Unlike the study here, Subramaniam et al. observed no above chance memory for stimuli presented in RSVP sequences at 72 or 128ms. Subramaniam et al. thus speculate that neuronal activity longer than 128ms is necessary to create a memory

trace for the occurrence of a stimulus. Based on this line of thinking, the memory performance in the task here should be null for brief SOAs (<126ms), as it was in Subramaniam et al.'s task for SOA as long as 126ms/image. This was not the case. Memory performance in our experiment was above chance at all rates, even at 14ms/image, when neuronal activity in STSa was curtailed to 71ms (see Chapter 7). It therefore appears that Subramaniam et al.'s (2000) hypothesis is probably wrong. On the other hand, our data is compatible with the idea that the stronger a neural response, the stronger the memory trace it will create, causing the reduced memory performance compared to the detection performance. The fact that Subramaniam et al. (2000) did not find evidence for stimulus memory at SOA where we found above chance performance may be due to the fact that we used relatively short RSVP sequences (7 images) compared to their longer sequences (32 images)[14], or to the fact that we used photographs, while they used line drawings.

A fourth possibility is based on **spatial attention**. In the detection task, the subject sees the target before the trial, and can direct his gaze and attention towards a point on the image that contains a distinctive feature that makes the image particularly easy to detect. The subjects attention can then remain in that position during the RSVP sequence, and rather than matching the entire complex image, he can then perform the task as a simple feature detection task in that particular location – given the location on which he focused his attention was well chosen, and contains a certain feature only if the target was present in the sequence. In the memory task, not knowing which of the 23 possible targets will be probed later on, the subject has no particular location on the image to focus upon, and will thus have to perform a more complex true pattern-matching task. Spatial attention can greatly benefit the processing of a particular location of a crowded scene, and it thought to do so by biasing the competition between the different items in favour of those in the target location (Duncan, 1998; Desimone 1998). This process could explain the difference between the detection and the memory task in the paradigm reported here.

---

[14] Note that in the electrophysiology we also used longer sequences, a problem addressed in section 9.3.3. 'Some words of caussion'.

9.4.2. Reaction time differences between the tasks

An interesting but curious effect is the fact that RT is ~250ms longer for the memory task than the detection task (the RT lines for the detection and memory task are parallel but offset in Figure 9.2). If task difficulty was to explain this effect, two conditions performed with the same accuracy should have the same RT. This is not the case: m111 and d28 have the same accuracy, but different RT (792ms and 532ms respectively, see Figure 9.2). Yet, next to task difficulty, the two tasks had an additional systematic difference that may explain the constant RT difference. In the detection task, the decision can be initiated after the occurrence of the target in the RSVP sequence. In that case the target is masked, curtailing the availability of information about the stimulus in the neural responses to a duration of SOA+60ms. Hence, the subject has nothing to gain from waiting longer before taking his decision, and takes it quickly. In the memory condition, the decision can only be initiated after the occurrence of the unmasked target at the end of the RSVP sequence. This stimulus is long (300ms) and unmasked. Hence, the subject is encouraged to take his time, for information regarding the stimulus will continue to rise with time.

9.4.3. Perception of brief stimuli – a critical discussion

A few words of caution are important in the interpretation of this data. Generalising from our finding, to saying that any image can be processed at 14ms/image in any sequence would be premature. It may be tempting to assert: "a single frame of advertisement in a cinema motion picture projection (42ms/image) can be perceived and memorised!". While the 73% correct detection and 62% correct memory accuracy at 42ms/image seems to support this claim, a few parameters of the RSVP paradigm used here are quite different from a motion picture situation. First, in motion pictures, a stream of *related* image is presented: subsequent snapshots of a car driving through the screen for instance. It is only at the relatively infrequent transitions between two sequences (e.g. when switching between two camera "shots") that unrelated images occur in direct succession. Our visual system has evolved to deal with continuous viewing conditions, and is likely to have evolved mechanisms that favour the processing of expected images that belong to the meaningful flow of images in a movie over the processing of intruding, unrelated and thus unexpected images. If a single frame, unrelated to other frames, is inserted under such conditions,

the single "odd-frame-out" is likely to suffer from the negative effects of selective attention being focused on the ongoing movie (see for instance Simons, 2000 for a review of evidences of "inattentional blindness", showing that subjects often miss salient features, if their attention is directed on a task that makes those features irrelevant). Detection performance under such conditions may be reduced. In addition, in a movie, much longer sequence of images are presented than in our RSVP sequences, and this is likely to result in poorer memory for individual frames if memory is tested for instance 1h after presentation. Experiments directly tailored to the question of perception of individual frames in movies are necessary to test perception and memory under motion picture projection conditions.

We decided to use sequences of only 7 stimuli in Experiment 2 because in each case we had to use the 7 distracters and the 1 target dictated to us by the single cells in the neurophysiological experiments. If we had repeated the sequence over and over again, as in the physiological experiments, it would have been difficult to measure the perception for individual frames: If the question had been "did you see this image in the sequence?", it would have been sufficient to perceive and memorise one of the occurrences of the image in the repeated sequences, and we would have received no direct information about the perception of each particular occurrence of the image in the sequence. If the task had been "how often did you see this image?", the problem of repetition blindness (e.g. Kanwisher, 1987) would have been introduced.

Repetition blindness refers to the fact, that if a stimulus is repeated in an RSVP sequence, the second occurrence of the same stimulus is often missed, even at between target intervals beyond those at which AB occurs. The second occurrence is probably not left unprocessed, but the percept caused by the second occurrence is not properly segregated from the percept caused by the first occurrence and is falsely interpreted as reflecting the persisting, remaining activity caused by the first experience. Repetition blindness is not surprising in the frame of persistence discussed in Chapter 3 and demonstrated in Experiment 1: activity caused by a stimulus outlast the offset of the stimulus by 60ms under severe masking conditions, and much longer under unmasked conditions (as will be shown later). Hence the visual system may assimilate prolonged neural activity not to the repeated occurrence of a stimulus but to the persistence of the representation of an earlier stimulus. Indeed, this is why we introduced a 500ms gap between the sequence and the target. At

shorter gaps, subjects clearly remember having seen the target, but were unsure whether it was actually in the sequence, or outside of the sequence (i.e. the target definition before or after the sequence).

In addition both in the physiological experiments and the psychophysical experiments, the same images are presented over an over again. Repeated presentation of an image is known to produce visual repetition priming (e.g. Subramaniam et al., 2000). It is unclear if the high detection performance observed in our experiments would also have occurred in the absence of such priming, if a new stimulus had been used on each trial. It will remain for future experiments to resolve that question.

9.4.4. Is psychophysical performance a measure of consciousness?

At the start of this chapter the following question was raised: "could they [the monkeys] perceive the stimuli their single [STSa] cells responded to, or did we measure subliminal neuronal activity?"

The results presented in this chapter indicate that at all presentation rates humans can perform above chance in a memory and a detection task. This in itself is unfortunately not a direct answer to this initial question of awareness. First humans rather than monkeys were used to measure the behavioural performance, but as outlined in the introduction, there is little reason to assume that monkeys would perform much differently in the same task (Leopold and Logothetis, 1996; Sheinberg and Logothetis, 1997; Logothetis, 1998; Kovács , 1995; Bridgeman, 1980; Fabre-Thorpe et al., 1998a,b).

Secondly, above chance performance in a psychophysical task demonstrates some form of perception of the stimuli, but not necessarily visual awareness of the stimuli: perception can be subliminal.

The difference between visually guided performance and conscious perception is particularly clear in pathological conditions such as blindsight. Blindsight patients have suffered V1 lesions and report that they are not aware of visual stimuli in the affected areas of their visual field. Nevertheless, such patients perform above chance on psychophysical tasks based on stimuli presented in their 'blind' field, while the patients explicitly report that they do not consciously perceive the stimuli: when asked

Box 9.1: The Mind-Brain problem and Empirical Science

The mind-brain problem refers to the question of how brain activity relate to the subjective experience we call the mind. Cognitive neuroscientists are fascinated by this problem, and it is important to understand what empirical science can contribute to this question and what it cannot.

The empirical scientific method is based on "inter-subjectivity", i.e. any empirical evidence must – in principal at least - be observable by any qualified observer. Behaviour fulfils this requirement, and so does brain activity, with the use of adequate methods, such as oscilloscopes etc., which convert single cell responses into an inter-subjectively observable quantity. Consciousness, and subjective awareness, *by definition* are only observable for the experiencing subject, which is why they are called "subjective awareness". As such they are *by definition* beyond the scope of the empirical scientific method. We all experience subjectivity, but we will never know empirically if any one else has the same experience or whether they just act as though they do.

Descartes pointed out very early that behaviour is a problematic indicator of consciousness. "Et je m'étois ici particulièrement arrêté à faire voir que s'il y avoit de telles machines qui eussent les organes et la figure extérieure d'un singe ou de quelque autre animal sans raison, nous n'aurions aucun moyen pour reconnoître qu'elles ne seroient pas en tout de même nature que ces animaux" (Renee Descartes, Discours de la méthode, In: *Oeuvres de Descartes*, publiées par victor cousin, Tome Premier, Paris, 1824 p185-186). Observable performance is not a proof of consciousness, and a machine that would press a key in the psychophysical experiment presented in this thesis with the same accuracy that a human, but without any awareness would be undistinguishable from a human, or a monkey. While Descartes believed that language could indicate consciousness, from the point of view of empirical science, the answer is simple and clear: no, it cannot. (See also 9.4.4.)

This does not mean, that the mind-brain problem should not be in the focus of scientific research, but one should be aware of what an empirical scientist can know, and what he/she will never know. Empirical science can investigate the relationship between observable brain activity, and observable *indicators* of subjective perception. Behavioural performance in a psychophysical task is such an *indicator* of conscious perception, in that we *believe* that to perform the task, conscious perception is used. What empirical science cannot know is whether these indicators really indicate conscious perception. Science stops, and belief starts when one goes from *indication* of consciousness to consciousness *per se*.

about attributes of stimuli in their 'blind' field, they report that they are just guessing (Weiskrantz, 1996; Stoerig & Cowey, 1997). In the present experiments, at the most rapid presentation rates, subject also report that they are often just guessing, and the example of blindsight patients illustrates, that the above chance psychophysical performance in the psychophysical tasks of this thesis is not a proof of a conscious perception of the stimuli.

Also in normal subjects, stimuli that are not consciously perceived can affect behaviour. This has been demonstrated in many priming experiments (e.g. Holender, 1986; Merikle, 1992; Dehaene et al., 1998). Dehaene et al. (1998) provide a particularly striking demonstration of this phenomenon. Subjects were presented with a sequence of 4 stimuli: a mask, a prime, a mask and a target. Subjects were asked to press a button with one hand, if the target stimulus represented a number smaller than 5, and another button with the other hand if the number was larger than 5. The number could be presented as a single Arabic digit (e.g. "2") or as a word ("TWO"). Unbeknown to the subjects, the prime, that was too short to be perceived, was also a number too (a digit or a word). In a separate experiment, half the primes were numbers, and half were non-words, and subjects were unable to indicate above chance if the prime was a word or a non-word, confirming that the primes were to short to be consciously perceived. Despite this fact, in their main experiment Dehaene et al. could demonstrate a classical priming effect: reaction times to decide if the target was above or below five were shorter if the subliminal prime and supraliminal target were congruent (i.e. both prime and target larger or both target and prime smaller than 5) than if the target and prime were incongruent (one smaller one larger than 5). This priming effect, occurring without visual awareness of the prime identity was also semantic. Even if the target was "ONE" and the prime was "3", a priming effect could be observed. In that case, priming can only occur if the subliminal prime was classified as "below 5" according to the instructions of the task. Hence a subliminal prime not accessible to consciousness can be successfully processed according to the requirements of the task, and influence behaviour. The authors than recorded the electroencephalogram (EEG) during the task, and could show that the prime created a deflection of the EEG that depended on the category of the prime (<5 vs >5), and was analogous to the deflection occurring before the overt motor response to a target of the same category, indicating that the prime was processed all the way to the preparation of the motor response after categorisation. fMRI confirmed these results, revealing an activation of motor cortex depending on the congruence/incongruence of the prime-target combination. Hence, even in normal subject, the entire response loop from visual stimulus to motor response in a psychophysical task can occur without conscious awareness of stimulus identity.

Indeed, also other neurophysiological experiments have demonstrated significant visual brain activity to stimuli that are not consciously perceived (Sahraie

et al., 1997; Stoerig et al., 1998, Wahlen et al., 1998; Morris et al., 1998; Berns et al., 1997), indicating that stimuli that are not consciously perceived can nevertheless be processed to a substantial extend by the brain without creating visual awareness of the identity of the stimulus. Hence, neither brain activity nor above chance performance in a psychophysical performance are necessarily indicators of conscious perception. Additional experiments will be needed to explore if the above chance performance observed in the most rapid presentation conditions in the present thesis represent above chance performance in the absence or the presence of conscious perception.

In addition, as pointed out in Box 9.1, empirical science cannot address directly the question of whether a stimulus is truly consciously perceived. Empirical science can only address the question of whether subjects report that they consciously perceived a stimulus - a slightly but importantly different issue. This distinction is particularly clear for RSVP stimuli, where the memory for stimuli appears to be particularly affected (e.g. Subramaniam et al., 2000): a stimulus could be consciously perceived, but by the time the subject is ask to report if he perceived the stimulus consciously, the subject may already have forgotten that he perceived the stimulus.

Yet, rather than focusing on what we do not know, let us focus on what we do know, namely that human performance reflects some knowledge of the identity of a stimulus in 14ms/image RSVP sequences. Where the term "perception" or "awareness" will be used, it will be in an operational fashion, as measured by the performance in a task. While I personally believe, that such performance is indeed correlated with subjective conscious experience of the stimuli, I am well aware, that there is and probably never will be proof to support such a claim, and hence I will refrain from making it. Indeed, I believe that as a continuous variable such as SOA is manipulated, the notion of subliminal vs. conscious is probably not a dichotomy, but rather a continuum, with clear perception and perfect performance on one side, and chance performance without any perception at the other side. In the middle, I believe perception to become less and less strong, fading progressively and gracefully. At very short SOA, this degraded perception is so weak compared to our normal experience, that we dismiss it as being conscious, while indeed, it is just a difference of quantity and not of quality. Indeed, our brain has to function with substantial amounts of spontaneous activity. The activity of a particular neurone, as a stimulus is briefly presented, will be barely larger than the spontaneous activity. Indeed, the

response distributions will overlap, with certain instances of the stimulus causing smaller activity than certain instances of spontaneous activity. Hence, to protect the visual system from permanent "false alerts", were spontaneous activity would be interpreted as stimulus evidence, the brain must be conservative in interpreting sensory evidence. Very brief stimuli may cause small neural activity which can affect the motor units pressing the response button, while the same neural activity will be so close to spontaneous activity that our consciousness may treat the neural activity as spontaneous.

At present it is nevertheless unclear if performance and conscious perception brake down jointly and gracefully as a single entity or whether perception is independent from performance and brakes down at a different SOA.

9.5. Summary and Conclusions

Single frames in RSVP sequences can be both memorised and detected at SOAs as short as 14ms/image. Hence, the performance observed electrophysiologically in Chapter 7 and 8 is not a spurious finding restricted to single cells in the rhesus monkey brain: perception and memory can occur for humans with the same stimuli at the same presentation durations. It remains unclear, whether the above chance performance of human subjects in Experiment 2 was always based on *conscious* perception of the stimuli: it may have occurred without conscious awareness of the stimuli. This performance nevertheless begs the question of how the cellular performance compares to the perceptual performance of human observers: a question addressed in the next chapter.

10. Single Cells and Perception I: RSVP sequences without gaps.

10.1. Introduction

In Chapter 7, single cell responses in the rhesus monkey STSa were shown to significantly discriminate between stimuli in RSVP sequences at 14ms/image. In Chapter 9, human observers were shown to have the capacity to do the same. In the present Chapter, in analogy to the work on motion detection by Newsome and colleagues (Britten et al., 1992, 1996; Celebrini and Newsome, 1994, 1995; Newsome et al., 1986, 1989; Salzman and Newsome, 1994; Shadlen et al., 1996) the relationship between single cell responses in the STSa and human psychophysical performance will be investigated.

The motivation for this comparison stems on one hand from the findings of Rolls et al. (1994a,b, 1999) and Kovacz et al. (1995), showing that the response of single neurones in the temporal cortex follow, to some extent, the perceptual performance under masking conditions, and on the other hand from the findings in binocular rivalry (Sheinberg and Logothetis, 1997 and Logothetis, 1998) which indicate that unlike earlier visual areas (V1, V2, V4), the vast majority of single cell responses in IT and STSa follow the percept (see Chapter 2). Together, these two lines of evidence suggest the possibility that cells in STSa might be particularly suited to support psychophysical decisions about the identity of an image in RSVP sequences, and call for a quantitative investigation of this possibility.

In the psychophysical tasks of Experiment 2, subjects had to judge if a particular stimulus was present in an RSVP sequence. STSa neurone responding selectively to one such stimulus in the sequence may be suited to support such a psychophysical decision. The goal of the present chapter will be to quantitatively investigate *how* suitable STSa neurones are for performing such psychophysical task by directly comparing the psychophysical performance of human observers measured in Experiment 2 with the 'neurometrically' derived performance of the single neurone recorded in Experiment 1. The term 'neurometric' here refers to the measurement of

how well an ideal observer could solve a psychophysical task if he based his decision exclusively on the firing rate of a single neurone, and has been created in analogy to the term psychometric. Three different neurometric methods will be explored in this chapter.

A bold statement suggested from the work of Logothetis and co-workers (Sheinberg and Logothetis, 1997 and Logothetis, 1998) is that neuronal responses in temporal cortex, and in particular in STSa are the neural correlate of object perception. If so, one would expect the neurometric performance of neurones recorded in our Experiment 1 to mirror the perceptual performance in Experiment 1 and thus, as illustrated in Fig. 9.1, to be affected by changes in SOA in ways very similar to the performance of human observers measured in the memory and the detection task of Experiment 2.

Of course, a single cell is not responsible for conscious perception or performance in a psychophysical task: a population composed of many cells will be required. If the ability to perform a discrimination of a single cell is quantified in this and following chapters, and if this performance of a single cell is compared to the performance of a human observer, this is not done to suggest that this single cell is responsible for the performance. The comparison is done with the idea that if a particular cell is part of a population of cells that participate in the psychophysical task, the performance of that single cell should be affected by changes in the stimuli in ways similar to the psychophysical performance of human subjects.

10.2. Methods

Both the single cell data and the psychophysical data used in this chapter have been collected according to the methods described in previous chapters. The psychophysical data required no processing in addition to those performed in Chapter 9. The single cell data, on the other hand, had to be analysed further, as will be described bellow.

10.2.1. Identification of sections of the RSVP sequences used in the physiological recordings which are equivalent to those used in the psychophysical investigation

The single cell data was collected at a time in which we had not planed to compare single cell results with psychophysical results. The aim was to collect as many trials as possible in a minimum amount of time, and hence continuous sequences were presented for as long as the monkey fixated. After collecting the psychophysical data, to perform neurometric analyses, it becomes necessary to identify sub-sequences in the continuous sequences used in the electrophysiological experiments which are comparable to the 7-image sequences used in the psychophysical experiments. Figure 10.1 describes how this was achieved, and how spike-counts are derived for sections of the electrophysiological recordings that were equivalent to target present and target absent trials in the psychophysical task. Spike-counts were always collected starting at the individual cell's latency. As for the duration of the spike-counting, both variable window durations depending on the SOA (see Table 7.2 column 4) and a fixed duration (71ms, the shortest of all windows in Table 7.2) were used.

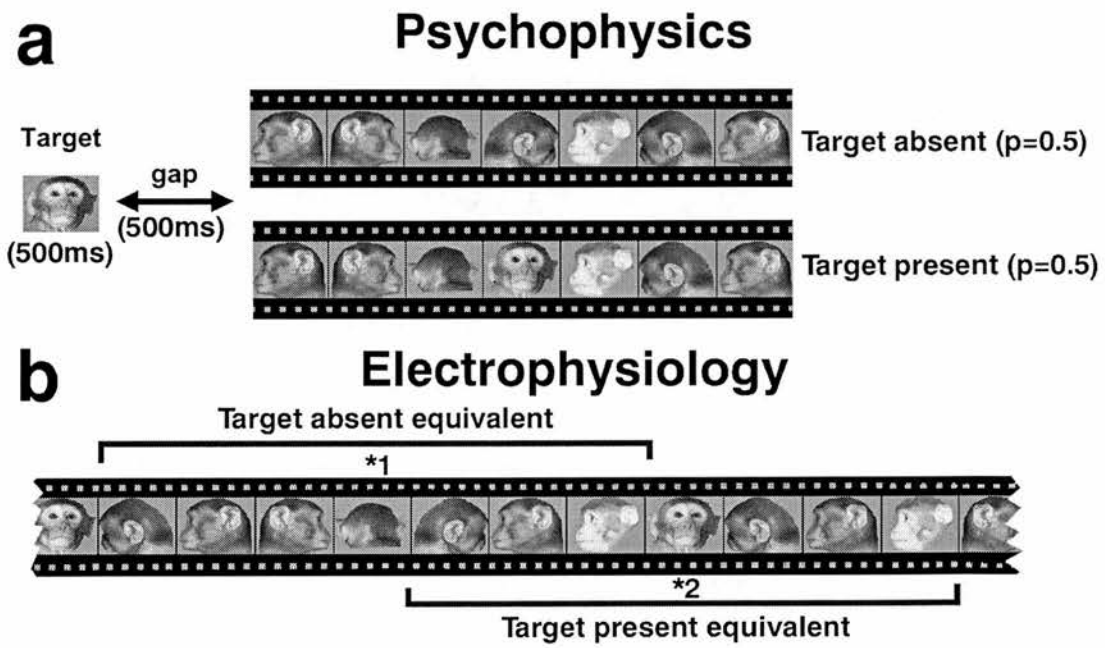


Figure 10.1: Identification of sequences in the electrophysiological record that are equivalent to sequences used in the psychophysical experiments. (a) The situation in the psychophysical experiments - here in the case of a detection task, but the same applies to the memory task. A target is to be detected in sequences of 7 items, that either contain exactly 1 (target present) or 0 (target absent) occurrences of the target in a central position. Each stimulus set used in the psychophysical experiment was derived from a particular neurone: the target corresponded to that neurone's best stimulus, and the 7 distracters corresponded to that neurones 'rest' stimuli (i.e. all but the best). The psychophysical results obtained for a certain stimulus set can thus be compared directly with the single cell results from a particular cell, using the same

stimuli. (b) Illustrates how within a continuous sequence used to stimulate the corresponding single cell (cell 154.2 in this case) sub-sequences can be found that are equivalent to the target present and target absent sequences used in the psychophysics. A 'target absent equivalent' subsequence is a sequence of images not containing at all the best or target stimulus. Spikes are then counted for the central stimulus within that sequence marked with *1. The time window in which the spikes are counted starts at the onset time of stimulus *1 plus the latency of the particular cell and lasts for either 71ms (fixed window) or the duration indicated in Table 7.2 column 4 (variable window). 'Target present equivalent' sequences contain the best stimulus in the central position and not in the flanking positions, and spike-counts for such sequences are taken relative to the best stimulus (*2) in the same way as for target absent sequences relative to *1. The exact number of rest stimuli required on either side of the *1 or *2 stimulus depend on the SOA. The rational behind not using a fixed number of flanking rest stimuli is explained in section 6.7.4 "Uncontaminated responses". This criterion is more lax at long SOA were trial numbers are small, and more stringent at short SOA were trial numbers are ample, while at the same time ensuring that the spike-counting window is not contaminated by neighbouring occurrences of the best stimulus - a situation that never occurred in the psychophysical task. This guaranties that the spike-counts reflect a situation similar to the psychophysical experiments.

10.2.2. 'Cellular decision' at optimal threshold method

The rational behind this method is simple. Since single cells responded most to the stimulus that was the 'target' in the psychophysical task, a single cell could be used to decide whether the target was present by simply applying a threshold rule. If the neurone showed a spike-count larger than a given threshold, an ideal observer would decide to respond that the target was present on this particular trial. If the spike-count was less than the threshold, the ideal observer would decide to answer that the target was absent from this trial. The observer is ideal, in that he knows which threshold is best for the task. Empirically, testing all possible thresholds, and selecting the one yielding the best performance replaces the ideal observer. Figure 10.2 illustrates this procedure on the example of a real cell at SOA=14ms/image. 'Cellular decision' refers to the target absent/target present response of the ideal observer based on the cellular response. Of course, this is a figure of speech, and a cell takes no decision.

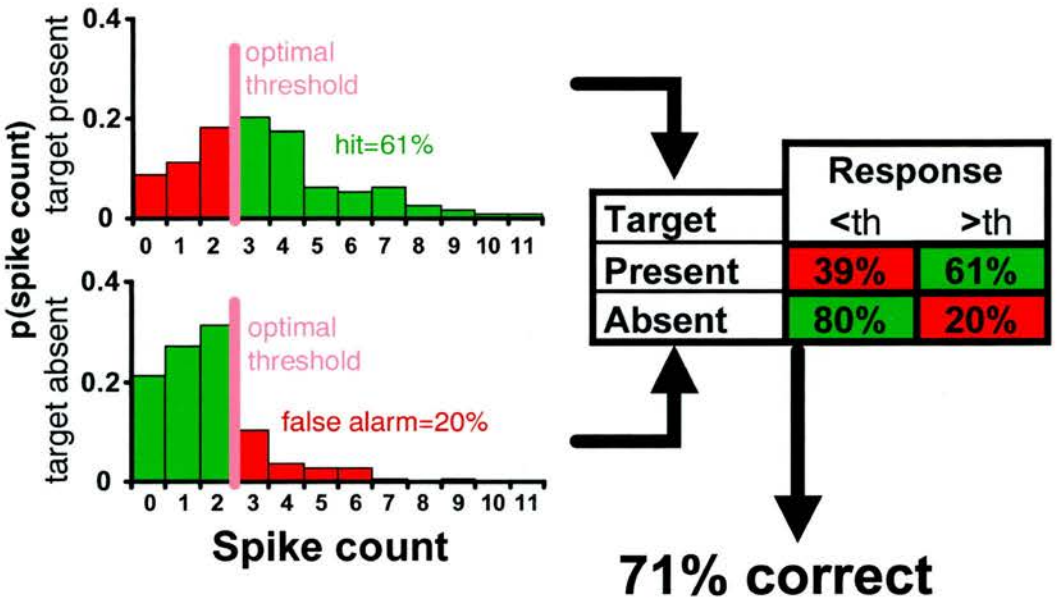


Figure 10.2: As illustrated in Fig. 10.1, spike-counts are taken for all target present and target absent sub-sequences of the single cell record for a particular cell (S143.9) at a particular SOA. These spike-counts are used to compile a spike-count histogram for target present and target absent sequences, with the x-axes representing the spike-count in the window, and the y-axis representing the probability of such spike-counts for that cell at that SOA. For any particular threshold (*th*; the vertical pink line in the histograms) a signal detection table comparable to that of Table 9.4 can be computed based on the threshold rule: respond 'present!' if the spike-count is larger than *th* and respond 'absent!' if the spike-count is less than *th*. 'Hit' trials are then trials in which the target is present and the spike-count is larger than *th*; 'miss' trials those where the target was present but the spike-count<*th*; 'false alarm' trials those in which the target was absent but spike-count>*th*; and 'correct rejection' trials those finally in which the target was absent and spike-count<*th*. Hit and correct rejection trials are correct responses and are indicated in green in the histograms and the table, while false alarm and miss trials are incorrect responses, and are indicated in red. A proportion correct responses can then be calculated as described in Equation 9.1 and was 71% correct for this particular cell at 14ms/image, for the optimal threshold. The optimal threshold is obtained by calculating the proportion correct cellular decisions for all possible thresholds and selecting the one yielding the highest proportion correct.

10.2.3. Receiver Operating Characterisic (ROC) analysis

While simple and intuitive, the optimal threshold analysis method described in the preceding section is based on an ideal observer's knowledge of the best threshold. To avoid this requirement, other authors (in particular see Newsome, Britten, Movshon and Shalden, 1989) have applied a Receiver Operating Characterisic (ROC) analysis of the neural data to yield a proportion correct response that does not depend on a particular threshold. This approach is described in Figure 10.3. While more

abstract, the surface under the ROC curve is the integration of the results of all possible thresholds, and is thus independent of the choice of a particular threshold. As can be seen in Figure 10.3e the results obtained using this method are very similar to those obtained using the optimal threshold method (if the ROC surface is correlated with the proportion correct obtained using the optimal threshold procedure for all cells at all SOA, $r^2=0.94$ for $n=186$ pairs and this correlation is highly significant: $p<0.01$). Indeed, if only one point from the ROC curve is known, as is the case for our psychophysical task, the two methods are mathematically equivalent, as demonstrated in Box 10.1.

10.2.4. Mutual information between stimulus and cellular decision at optimal threshold

In Chapter 9 an information theoretical analysis of the signal detection table of the human observers has been performed. This information calculated in Chapter 9 is the information between the response of the observer and the stimulus. This information is different from the information between the neural firing rate and the stimulus calculated in Chapter 8, in that it calculates the information between the stimulus and the *decision* rather than the spike-count this decision is based upon. The decision of the ideal observer of section 10.2.2. is a "present/absent" decision which unlike the firing rate of the neurone can be directly compared with the information calculated in Chapter 9. The mutual information between the cellular decision at optimal threshold and the stimulus is calculated simply by applying the same Equation 9.2. of section 9.5.2.2. to the signal detection tables obtained at optimal threshold for each cell and SOA as described in Figure 10.2. Note that in this case, optimal threshold refers to threshold with the highest information rather than the threshold with the highest proportion correct.

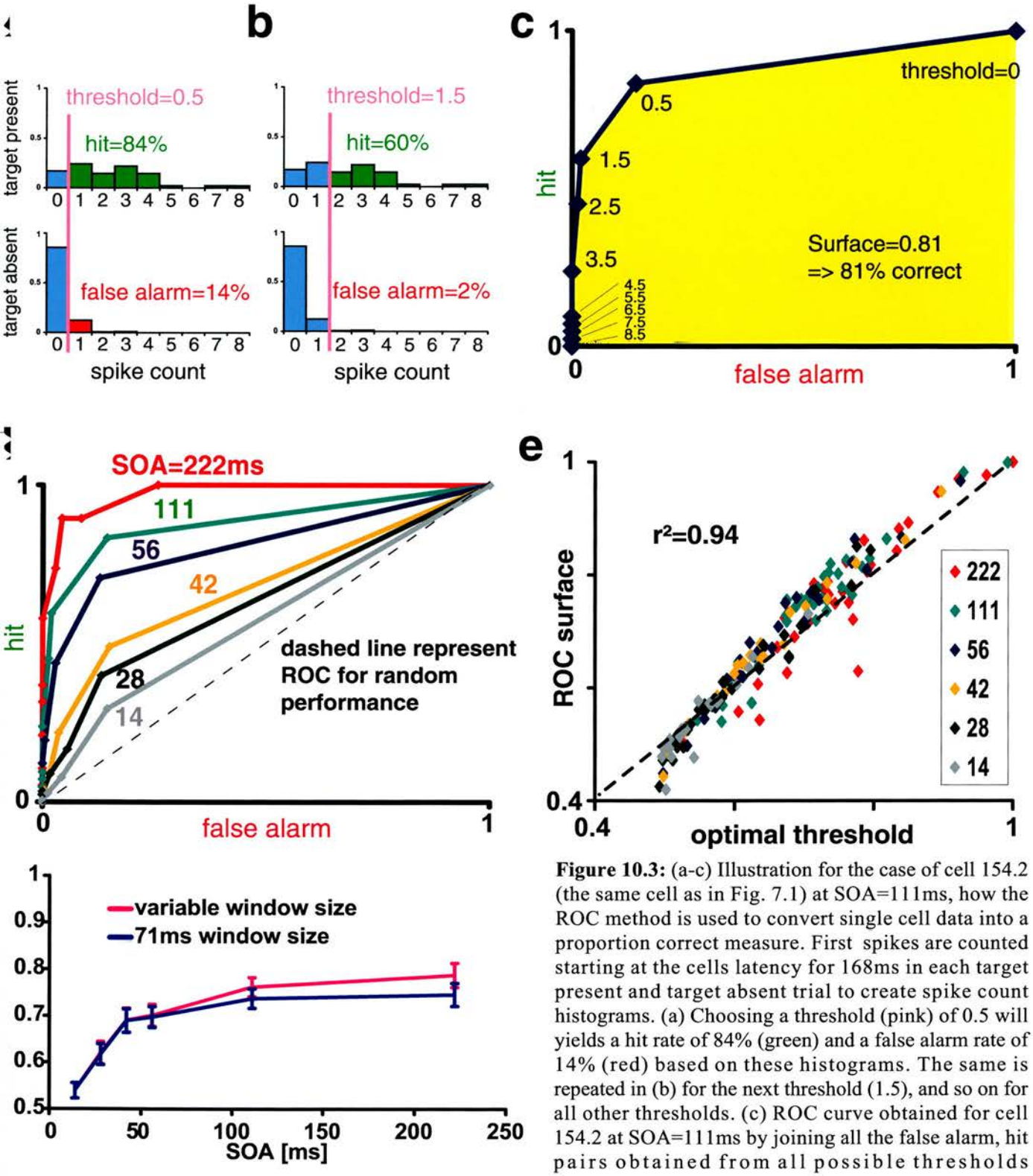
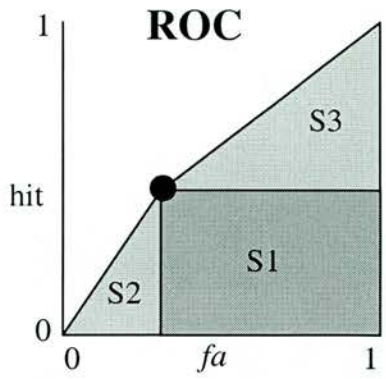


Figure 10.3: (a-c) Illustration for the case of cell 154.2 (the same cell as in Fig. 7.1) at SOA=111ms, how the ROC method is used to convert single cell data into a proportion correct measure. First spikes are counted starting at the cells latency for 168ms in each target present and target absent trial to create spike count histograms. (a) Choosing a threshold (pink) of 0.5 will yields a hit rate of 84% (green) and a false alarm rate of 14% (red) based on these histograms. The same is repeated in (b) for the next threshold (1.5), and so on for all other thresholds. (c) ROC curve obtained for cell 154.2 at SOA=111ms by joining all the false alarm, hit pairs obtained from all possible thresholds rresponding thresholds are indicated). Proportion correct is measured as the (yellow) surface under the curve. (d) ROC ves for cell 154.2 at different SOA. When SOA is reduced, the ROC curves get closer and closer to the dotted diagonal, icating decreased performance, but all point are above the dotted diagonal, indicating that the cell would be adequate for ve chance performance in a psychophysical task whatever threshold is chosen. (e) ROC surface (y-axis) is very comparable =0.94) to the proportion correct obtained using the optimal threshold method described before (x-axes). Each point is one at a given SOA. In most of the cases, the ROC surface is slightly larger than the prop, correct at optimal threshold, as icated by the fact that most points lye above the dashed diagonal (see also Box 10.1). (f) The mean (±sem) of the ROC ace calculated for 22 cells (18 cells for 42ms/image) as a function of SOA. The blue line represents the value obtained if e counts are taken in the shortest of all windows (71ms) starting at each cells latency. The pink line represents the very ilar results obtained if the duration indicated in Table 7.2 column 4 is used, which is longer for longer SOA.

Proportion correct decisions $\overset{?}{=}$



$$
\begin{aligned}
\Leftrightarrow \quad & \textit{proportion correct} && = && \textit{surface under ROC curve} \\
\Leftrightarrow \quad & \textit{proportion correct} && = && S1 + S2 + S3 \\
\Leftrightarrow \quad & \frac{\textit{hit} + \textit{correct rejection}}{\textit{hit} + \textit{fa} + \textit{correct rej} + \textit{miss}} && = && \textit{hit} \cdot (1 - \textit{fa}) + \frac{\textit{hit} \cdot \textit{fa}}{2} + \frac{(1 - \textit{hit})(1 - \textit{fa})}{2} \\
\Leftrightarrow \quad & \frac{\textit{hit} + (1 - \textit{fa})}{2} && = && \frac{2\textit{hit}(1 - \textit{fa})}{2} + \frac{\textit{hit} \cdot \textit{fa}}{2} + \frac{(1 - \textit{hit})(1 - \textit{fa})}{2} \\
\Leftrightarrow \quad & \textit{hit} + 1 - \textit{fa} && = && 2\textit{hit} - 2\textit{hit} \cdot \textit{fa} + \textit{hit} \cdot \textit{fa} + 1 - \textit{fa} - \textit{hit} + \textit{hit} \cdot \textit{fa} \\
\Leftrightarrow \quad & 1 + \textit{hit} - \textit{fa} && = && 1 + \textit{hit} - \textit{fa} \\
\Leftrightarrow \quad & 0 && = && 0
\end{aligned}
$$

Box 10.1: How does the proportion correct decision compare to the surface under the ROC curve? This box illustrates how, if only a single data point (black circle on the right ROC plot) on the ROC curve is known, the two methods are mathematically identical. This is for instance the case for our psychophysical task, where subjects were free to choose their own decision criterion, and were not encouraged to change it. On the right of the equal sign, the ROC surface is calculated, based on the fact that the ROC curve always stats at (0,0) and finishes at (1,1) and has to pass through the single data point. *Hit + miss* always adds to one, and so does *fa* (false alarm) and *correct rejections*. If more than one point is known, the surface tends to be larger, but can also be smaller than the proportion correct, as illustrated in Figure 10.3e.

10.3. Results

The proportion correct 'cellular decisions' were calculated using both the optimal criterion and the ROC surface analysis for 22 cells that had been tested at 14ms/image using the same stimulus sets as those used in Experiment 2 (a 23rd cell, also tested at 14ms/image was the reason for testing 23 stimulus sets in Experiment 2. Unfortunately, for this cell, the detailed sequence of stimuli used was lost, and both the data from the single cell, and the corresponding stimulus set in the psychophysical experiment will be excluded from this chapter). First, an analysis based on spike-counts in the entire time window for response analysis was performed to reflect the best performance possible based on the single cell. The analysis was repeated using a constant window of analysis of 71ms to ensure that the observed effects are not due to using longer spike-counting windows at longer SOA. Additionally, an information theoretical analysis of the data was performed.

10.3.1. Proportion correct cellular decisions for single cells using the optimal threshold method.

As can be seen in Figure 10.4a, the cellular decision of individual STSa cells becomes increasingly accurate with increasing SOA. An ANOVA was performed comparing the 22 cells for 6 SOA x 2 windows (the variable window and the 71ms window) in a within subject design. If all 6 SOA, including the 42ms/image condition for which only 18 of the cells were tested, there was no main effect for window of analysis ($F(1,17)=1.17$, $p>0.29$) and no SOA x Window interaction ($F(5,85)=1.75$, $p>0.13$), indicating that considering spikes beyond 71ms after the cells response onset latency did not have much effect on the accuracy of the cellular decision. There was a highly significant effect of SOA ($F(5,85)=42$, $p<0.0001$), with a Newman-Keuls post-hoc test[15] indicating that all pairs except for 42ms/image vs. 56 ms/image and 111ms/image vs. 222ms/image were significantly different. If the 42ms/image condition was excluded from analysis to increase n to 22 cells, the effect of window of analysis became more reliable: the main effect of window almost reaching significance ($F(1,21)=3.82$, $p<0.065$), and the SOA x Window interaction becoming significant ($F(4,84)=3.9$, $p<0.01$), indicating that for longer SOA, the window of

[15] All Newman-Keuls post-hoc tests were performed using a critical range of alpha=0.05

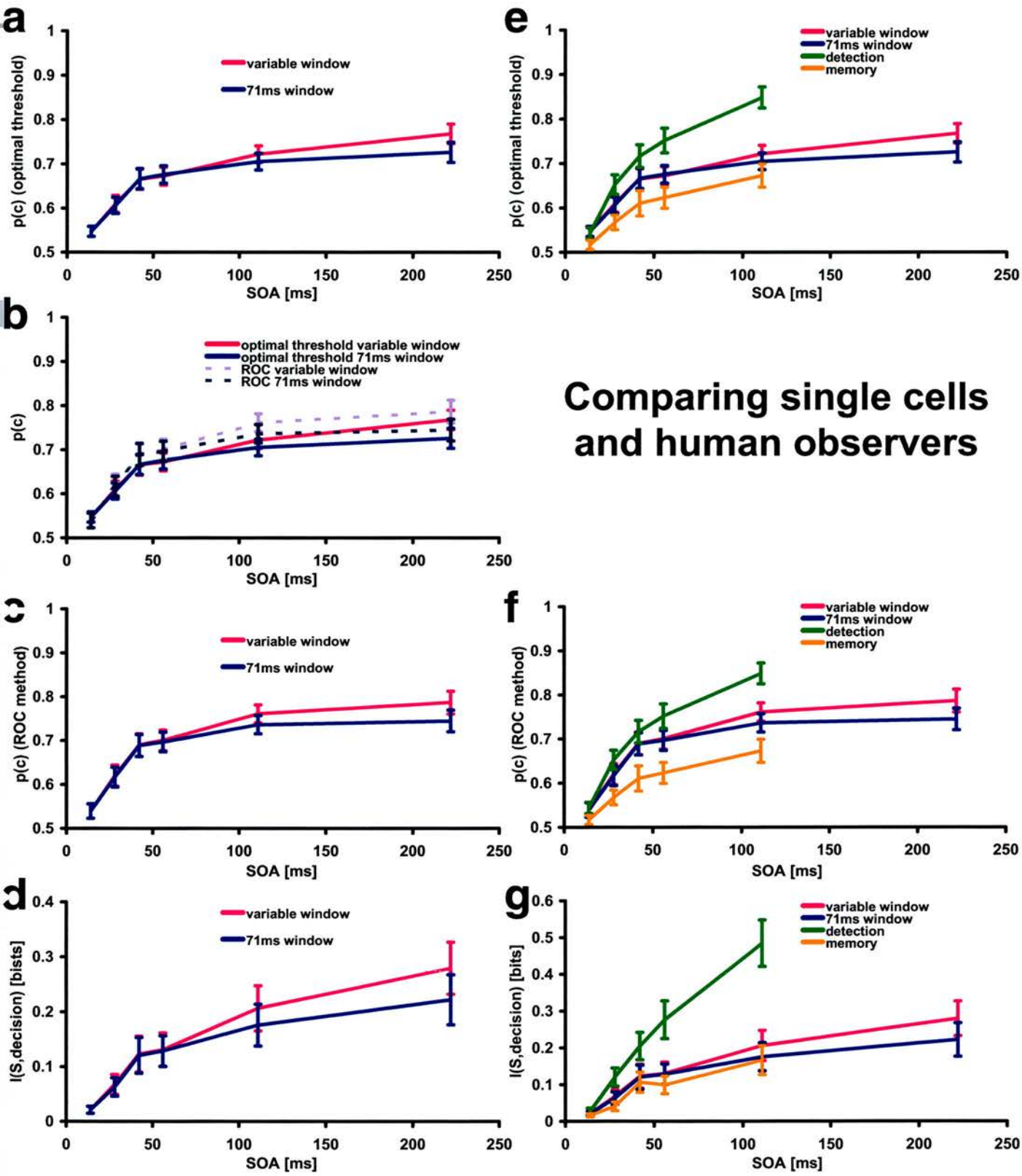


Figure 10.4: (a-d) Results of the neurometric analyses. (e-g) Compares the neurometric performance with the sychophysical performance. The same colour code is kept throughout the graph. (a) Proportion correct [p(c)] ellular decisions at optimal threshold for two spike-counting windows. (c) Proportion correct cellular decisions stimated by the surface under the ROC curve. (b) Compares the results from the optimal threshold (solid lines) and ıe ROC method (dotted lines), showing how similar they are, except for the slightly larger values obtained in the :OC method. (d) Mutual information in bits between the stimulus presence and the cellular decision reached at ptimal threshold. (e,f) Same as (a,c) respectively but together with the proportion correct decisions achieved by 5 uman observers in the memory and the detection task (see Chapter 9, Fig. 9.2). (g) Same as (d) together with the ıutual information between the decisions reached by the 5 human observers and the presence of the stimulus. Note ow, whatever neurometric measure is used, the single cell performance is always bracketed by the single cell data. =22 (cells or stimulus sets), for all conditions, except for the single cell data at 42ms/image, where n=18.

analysis has more importance (which is not surprising, since the variable window becomes increasingly longer as SOA is increased, being also 71ms at SOA=14ms/image but 284ms long at 222ms/image). Indeed, a Newman-Keuls post-hoc analysis[1] revealed that only in the 222ms/image condition did the two windows differ.

Accuracy increased when SOA was prolonged from 14ms/image to 56ms/image. Thereafter, accuracy increases only modestly, with no significant difference between 111ms/image and 222ms/image. Accuracy was above chance (>0.5) at all SOA, even at 14ms/image ($t(21)$=(mean-0.5)/s.e.m=4.1, $p<0.001$, two tailed).

10.3.2. ROC surface analysis.

Figure 10.4c illustrates the proportion correct response derived for the 22 (or 18 at 42ms/image) cells included in the analysis.

The findings are virtually identical to those obtained using the optimal threshold method. An ANOVA with 2 windows x 6 SOA (including 42ms/image) indicated a main effect of SOA ($F(5,85)=41$, $p<0.0001$), with a Newman-Keuls[1] indicating that all SOA pairs are different except the 111ms/image vs. 222ms/image. There was no main effect of window ($F(1,17)=2.4$, $p>0.14$) and no SOA x window interaction ($F(5,85)=1.8$, $p>0.12$). If the 42ms/image condition was excluded to increase n to 22, again, window gained in importance, with its main effect approaching significance ($F(1,21)=4.12$, $p>0.055$), and the interaction reaching significance ($F(4,84)=2.8$, $p<0.04$), with a Newman-Keuls[1] revealing that only in the 222ms/image condition did the two windows differ.

Again, as for the optimal threshold analysis, cellular decision accuracy was above chance (>0.5) at all SOA, even at 14ms/image ($t(21)$=(mean-0.5)/s.e.m=2.4, $p<0.03$, two tailed).

10.3.3. Comparing ROC with optimal threshold

The ROC and the optimal threshold methods yielded very similar results, as mentioned earlier in the methods section (Fig. 10.2, Box 10.1), with $r^2=0.94$. Fig. 10.4b illustrates the accuracy derived by the two methods as a function of SOA and window.

A 2 methods (ROC vs. optimal threshold) x 6 SOA (including 42ms/image) x

2 window within subject ANOVA revealed a significant main effect of analysis method ($F(1,17)=63$, $p<0.001$) and a method x SOA interaction ($F(5,85)=11$, $p<0.001$) (all other interaction $p>0.05$), with the ROC method yielding slightly larger proportions correct than the optimal threshold method, especially at long SOA. The same was true if the 42ms/image condition was excluded. A Newman-Keuls[1] revealed that only at 14ms/image was there no significant difference between the two methods. The larger difference between the two methods for larger SOA, is probably due to the fact, that as the window of analysis is increased, so is the maximum number of spikes within the window. The larger the maximum spike-count, the higher the number of thresholds that yield different results, and hence, the more points will be on the ROC curve. The more points there are on the ROC curve, the more the surface under the curve can deviate from the proportion correct (see Box 10.1).

10.3.5. Mutual information between stimulus and cellular decision at optimal threshold.

Figure 10.4d illustrates the mutual information between the cellular decision at optimal threshold and the presence/absence of the target. As for proportion correct, information increases with increasing SOA, especially as SOA is increased from 14ms/image to 56ms/image, and less steeply thereafter. A 2 windows x 6 SOA (including 42ms/image) ANOVA showed a significant main effect of SOA ($F(5,85)=16, p<0.0001$), with a Newman-Keuls[1] indicating that all but the 42ms/image vs. 56ms/image SOA pairs are significantly different. The ANOVA showed no main effect of window ($F(1,17)=2.75, p>0.11$), but an interaction with SOA ($F(5,85)=2.36$, $p<0.05$), and a Newman-Keuls indicated that only at 222ms/image did the two windows differ significantly. If the 42ms/image condition was excluded, the main effect of window became significant in addition to the other effects ($F(1,21)=4.97$, $p<0.04$), and a Newman-Keuls[1] indicates a difference between the two windows at both 111ms/image and 222ms/image.

10.3.6. Comparing the psychophysical results with the proportion correct cellular decisions at optimal criterion.

Fig 10.4e illustrates the relation between the proportion correct cellular decision at optimal threshold and the psychophysical proportion correct decisions in the memory and detection task. The cellular performance lies between the two

psychophysical curves. The precise relationship between cellular and human decisions were explored using a series of 2 way ANOVAs. SOA was always one of the two factors, and the analysis was run both with and without the 42ms/image condition. The second factor was arbitrarily called PN (i.e. Psychophysics vs. Neurometrics). Both the fixed and the variable window of analysis were analysed in this fashion. A main effect of PN or an SOAxPN interaction then indicates, that the accuracy of the cellular decisions differed from that of the human decisions.

Single cells vs. detection: A 2 PN (detection vs. single cells) x 5 SOA gave a significant main effect for PN for both windows ($F(1,17)=6.33$, $p<0.03$ for the variable, and $F(1,17)=6.82$, $p<0.02$ for the fixed window), indicating that human detection was more accurate than the single cell decision. There was also a significant interaction between SOA and PN ($F(4,68)=8.6$, $p<0.0001$ for variable and $f(4,68)=10.18$, $p<0.0001$ for fixed window), due to the fact that the difference between the single cells and the detection task increased with increasing SOA.

Single cells vs. memory: A 2 PN (memory vs. single cells) x 5 SOA gave no significant PN main effect ($F(1,17)=3.09$, $p>0.09$ for variable, and $F(1,17)=3.08$, $p>0.09$), nor significant interactions (PNxSOA), indicating that the single cell performance was too close to the memory task to show a significant difference. If SOA=42ms/image was excluded to increase n to 22 cells, the main effect of PN became significant at $p<0.05$, indicating that the single cell performance is indeed slightly better than the human performance in the memory task, but that the difference requires a large n to become significant. The interaction was never significant, indicating that human performance in the memory task is affected by SOA in the same way as the single cell performance.

10.3.7. Comparing the psychophysical results with the ROC surface derived proportion correct cellular decisions.

Figure 10.4f illustrates how the accuracy of the cellular decisions compared to that of the human decisions, if the ROC method was used to assess accuracy in both cases (Note that for the human observers, the results are identical using both methods, see Box 10.1). As can be seen, the ROC method yielded slightly larger results than the optimal threshold method, which increases the difference between single cells and the human observers in the memory task, while decreasing the difference in the detection

task. The same set of ANOVAs described in 10.3.6. was performed to test this observation.

Single cells vs. detection: Indeed the main effect of PN (detection vs. single cells) was not significant for both the variable and the fixed window if all SOAs were considered ($F(1,17)=2.6$, $p>0.1$ for variable, and $F(1,17)=3.4$, $p>0.08$), and even if the 42ms/image condition was excluded, the main effect remained non-significant for the variable window ($F(1,21)=4$, $p>0.059$), but reached significance for the fixed window ($F(1,21)=6$, $p<0.03$). The interaction between SOA and PN was always significant (all $p<0.05$) , and indicates that especially at large SOA, human observers in the detection task remained more accurate in their decisions that the single cells. (e.g. $F(4,68)=3$, $p<0.03$ for the variable window including 42ms/image).

Single cells vs. memory: The higher ROC performance resulted in a main effect ($p<0.05$) of PN in all 4 ANOVAs: for both fixed and variable window, including and excluding the 42ms/image condition (e.g. $F(1,17)=6$, $p<0.03$, for the variable window including 42ms/image). As for the optimal threshold method, the interaction never reached significance (all $p>0.06$), confirming that human performance in the memory task is affected by SOA in the same way as the single cell performance.

10.3.7. Comparing single cell and human performance based on information

Fig 10.4g illustrates how much information about the presence of the target was available from the decision of human observers, and from the decision of the ideal observer of the single cell responses. From that figure, it appears that the memory task but not the detection task yields information that is not different from that available from the cellular decisions at optimal threshold. Again the same ANOVAs were performed to test this observation.

Single cells vs. memory: In all 4 cases, i.e. for the fixed and the variable window, including or excluding the 42ms/image condition, there was never a significant main effect of PN or a significant interaction between PN and SOA (all $p>0.3$), indicating that indeed human subjects in the memory task, and single cells at optimal threshold reached decisions that were virtually equally informative about the presence of the stimulus in the sequence, and that both were equally effected by SOA.

Single cells vs. detection: in all 4 ANOVAs the main effect of PN was significant at $p<0.01$, and the interaction PN x SOA also ($p<0.01$), indicating that

while the memory task matched the single cell data, the human decisions in the detection task contained more information about the presence of the stimulus than evident in the cellular decisions.

10.3.8. Comparing the difficulty of a stimulus for humans and single cells

To investigate if the stimulus sets that yielded poor performance in the single cells were also causing poor performance in human observers, the accuracy scores obtained using the optimal threshold and the ROC method were correlated with each other and with the proportion correct in the psychophysical tasks. While the 2 single cell scores correlated highly with each other (e.g. ROC, variable window 28ms/image and optimal threshold fixed window 56ms/image, $r=0.88$, $p<0.001$), and the 2 psychophysical tasks correlated highly with each other (e.g. detection 28ms/image and memory 111ms/image, $r=0.74$, $p<0.001$), not a single correlation coefficient between a single cell performance and a psychophysical task was significant (all $p>0.05$), indicating that although overall performance is similar, a difficult stimulus for a human to detect is not a stimulus for which we found poor single cell discrimination.

The same was true if correlations were based on information instead of accuracy (i.e. proportion correct).

10.4. Discussion

Three neurometric analyses were used to derive measures of neuronal response that could be directly compared with the psychophysical performance of human observers: the proportion correct decisions at optimal threshold, the surface under the ROC curve, and the information contained in the decisions at optimal threshold. All three methods yielded basically identical results: the performance of single cells always fell in the range between the human performance in the memory and the detection task. The fact that three different methods yielded to the same result illustrates the robustness of this finding.

10.4.1. The single cell responses in STSa are suitable for solving the psychophysical task

The results of Experiment 2 are exactly what would be expected, if neurones in the STSa were directly involved in making the psychophysical decision. As illustrated in Figure 9.1, in the memory task, the target had to be perceived, and memorised. Hence the human performance in the memory task underestimates the processing that would occur under the physiological conditions, where no memory is required. The fact that single cell performance is higher than human performance in the memory task is compatible with the idea that a relatively small number of neurones may be sufficient to perform a psychophysical task. In addition, there was never a significant interaction with SOA if memory was compared with the performance of single neurones, illustrating that both were equally affected by SOA: again, exactly what would be expected, if the neurones in STSa directly participated in the performance of the memory task. Clearly, a single neurone will not be responsible for the perception or the psychophysical performance: a number of neurones will be necessary.

Due to the selective attention directed towards the target in the detection condition, one would expect better performance in that task compared to perception under the physiological conditions in which no systematic selective attention is directed towards a particular stimulus. Indeed, the literature on selective attention (in particular Treue and Martinex Trujillo, 1999), indicates that selective attention to a particular feature and/or location selectively increases the response of neurones representing this feature and location in the dorsal stream (MT). Such an increase in neural activity of neurones involved in the representation of the target stimulus would increase the accuracy of the cellular decisions in our task, and could account for the discrepancies in performance between the single cells recorded without selective attention, and the human performance in the detection task with selective attention.

Our single neurone data is therefore compatible with the idea, that the nervous system could base its decision regarding the presence or absence of a particular image in a sequence, upon the firing rate of a population of single neurones in the STSa, and by doing this perform as well as human observers. From our data alone, it is impossible to know exactly how many neurones would be required to perform such a task, but our data suggest that a relatively small number may be sufficient, given that

a single neurone's performance already falls on average between the human performance in the memory and the detection task.

10.4.2. Single cells in STSa and visual perception

As discussed in section 9.4.4. above chance performance in a psychophysical task does not necessarily show that subjects consciously perceive the stimuli they base their decisions upon. Depending on whether the psychophysical performance was or was not based on conscious perception, the implications of the present data are different.

If human subjects based their responses in the psychophysical tasks on their subjective percept of the target image in the sequence, it is tempting to believe, that single neurones in STSa may be responsible for creating the percept of the target image. This idea is particularly tempting considering that Logothetis and coworkers (see Chapter 2) showed that for binocular rivalry, the firing rate of neurones in areas in the ventral stream earlier than IT do not correlate with the reported percept. Later in this thesis, data will be presented that suggest, that STSa neurones may be responsible, not for all aspects of the perception of the target, but only for the perception of stimulus *identity*. In addition, it is tempting to suggest a quantitatively graded concept of conscious perception, in which strong firing of STSa neurones is associated with a vivid and robust percept, while a weaker firing is associated with a degraded mental image of the represented stimulus. This would explain why as SOA is decreased, the accuracy of the human decisions decreases together with the accuracy of the cellular responses (Figure 10.4e-g). Smaller responses may create poorer visual awareness of the stimulus and based on those poorer percepts, subjects would perform less accurately in the psychophysical tasks. As discussed earlier (Box 9.1 and Section 9.4.4), there is unfortunately no good way to ensure that psychophysical performance is indeed always based on a conscious percept, and these conclusions are therefore particularly tentative, and based on personal beliefs rather to scientific conclusions.

What can be said with certainty, is that the data of this Chapter is compatible with the idea that STSa may play a central role in the psychophysical decision of human subjects in Experiment 2. Whether consciousness for the identity of the stimuli occurs in the same process remains unknown.

10.4.3. Discussion of some limitations of the human vs. single cell comparison.

The two above statements, that single neurones in STSa are suitable for supporting psychophysical performance, and the fact that the single STSa cells may be involved in creating the percept of the target image in the sequence are to be made carefully. While our data supports especially the former claims, it by no means proves them. This is particularly so, because of a number of caveats of comparing of single cell performance in monkeys with psychophysical performance measured separately in humans. Some of these caveats will be discussed bellow.

10.4.3.1. The use of different species.

The psychophysical data was collected from humans, while the single cell data has been collected from monkeys. There is evidence discussed earlier, that the performance of the monkey and human visual system are extremely similar, especially under masking conditions [Chapter 9: "Both species respond in similar way to binocular rivalry (Leopold and Logothetis, 1996; Sheinberg and Logothetis, 1997; Logothetis, 1998; see section 2.2.4). With respect to the speed of visual processing in the two species Fabre-Thorpe et al. (1998a,b) could show that monkeys and humans take comparable times to decide if a visually presented scene contains an animal. Most importantly, in masking experiments, both species respond similarly to metacontrast masking (Bridgeman, 1980) and their performance at detecting a particular target when it is followed by a pattern mask is virtually identical (Kovács et al., 1995, results illustrated in our Fig. 3.17)"]. Indeed, other authors have performed a direct comparison between monkey single cell data and human psychophysical data under masking conditions (e.g. Rolls et al., 1994a, 1994b, 1999; Macknik and Livingstone, 1998).

Yet, it is not impossible that the perceptual system of monkeys may be slightly more rapid that that of human subjects. Thorpe and colleagues (Fabre-Thorpe et al., 1998a,b) measured the reaction time necessary for classifying natural, visually presented scenes in both species. While the reaction times they measured are similar in the two species (235ms for humans and 190ms for monkeys, if the first 10ms time bin at which hit trials significantly outnumber false alarm trials is taken), they nevertheless are slightly different, with monkeys showing faster performance. Those differences can be accounted for, at least in part, by differences in the processes occurring after perception (e.g. longer conduction time in the motor nerves

innervating the hand in humans due to the sheer length of the nerves). Indeed, while human's shortest correct RTs are 235ms long, evoked potentials recorded in humans show significant evidence of correct classification as early as 150ms after scene onset. Nevertheless: the difference between human and monkey RTs may also be due in part to slightly faster visual processing time in monkeys compared to humans.

Kovács et al. (1995) found identical performance for the two species for masked stimuli at a given SOA (see our Figure 3.17). When stimuli were unmasked, the monkey performance was less accurate than the human performance. In a personal communication he attributed that difference to the fact that the monkey's were not yet fully trained, and probably did not collaborate in the task as fully as the highly motivated human observers (including one author). If the monkeys did not fully collaborate in the unmasked condition, it is likely that they did not either fully collaborate in the masked condition. In that case, equal performance for humans and monkeys in the masked conditions may actually indicate that the monkeys were slightly better than humans, but just did not always report what they saw. Indeed, one can only ever know what a subject decides to report, and never what he perceived and this illustrates the fact that, even if monkeys had been trained to report their perception, this report still would have yielded *ambiguous* information regarding the monkeys perception.

What is important though, is that if measuring perception in humans instead of monkeys may indeed have *slightly* underestimated the perceptionnal efficiency possible in monkeys at equal SOA, this would not have changed the interpretation of our findings. The main finding of our experiment is the similarity in the way human and single cell performance are affected in similar ways by changes in SOA. We do not claim, that *one* single neurone is responsible for the psychophysical performance, but rather that *population* of neurones might be a sufficient basis for the observed human performance in a psychophysical decision. Such a statement would hold even if the psychophysical performance had been slightly superior if measured in monkeys. And from both the work of Kovács et al. (1995) and Fabre-Thorpe et al. (1998a,b) more than slight differences between the species are unlikely. Hence, there is reason to believe, that our conclusions would have been identical, if monkeys had been used as psychophysical observers rather than humans.

10.4.3.2. The lack of trial-by-trial comparison.

The psychophysical measures were not taken from the monkeys while the single cells were recorded. This makes it impossible to perform a trial-by-trial comparison of single cell performance and perceptual report. Trial-by-trial comparisons are powerful, because if a neurone is thought to be part of a population of neurones playing a central role in a psychophysical task, then trials on which the neurone responds less should be trials in which the monkeys is less likely to report target perception, and *vice versa*, trials in which the neurone responded more should be trials in which the monkey is more likely to report target perception, or reports target perception with greater certainty (if choice certainty is assessed in addition to a simple 2AFC). If we had found, that the firing of neurones in STSa had not followed the pattern of perception in humans, it would have told us, that it is probably not worth training monkeys to report their perception. Given the present results, it appears that investing the time necessary for such training would be worthwhile, and would allow a more stringent test of our hypothesis. Rather than giving definite evidence, the present experiments should thus be seen as a primer – a first quantitative evidence towards a link between STSa firing and perception of image identity, especially under RSVP conditions.

Even if such trial-by-trial perceptual report would have been available, as in the case of Newsome and colleagues' and Logothetis and coworkers' experiments, and even if STSa neurones' trial-by-trial responses had been found to correlate with the perceptual report, it would still not have been proof of the direct involvement of STSa cells in the psychophysical decision or in conscious perception. This is because, unless one can show that these neurones are the only neurones that correlate with the report, there is always the possibility, that other neurones are the ones being directly involved in the perceptual report. Logothetis and coworker (e.g. Logothetis, 1998) are in that problematic situation: while the activity of 90% of neurones in IT/STS is correlated with the reported percept during binocular rivalry, suggesting that IT and STS may be the site in which the percept is created, about 20% of neurones in V1/V2 also correlate with perception, making it impossible to state that only IT and STS would be suitable to create the percept of the images during binocular rivalry. Electrical micro-stimulations, such as those of Celebrini and Newsome (1995) are a precious further step in establishing such causal relationships, by showing that manipulating neural activity in the absence of stimulus manipulation can influence

performance. Nevertheless even knowing that electrical micro-stimulation in one brain area influences performance does not exclude the possibility that the electrically induced neuronal activity of this area influenced activity in another area, and that it was this second area and not the stimulated area that had the direct link to behaviour and perception.

The present findings should thus represent a primer to perform two experiments that could test the tentative conclusions of the present chapter: training monkeys to report their perception in RSVP sequences while single cells in STSa are recorded and applying electrical microstimulation in STSa while the monkeys report their perception, especially at short SOA, where perception is close to threshold.

10.4.3.3. Selecting stimuli based on the single cell responses

A further caveat of our experiment is the fact that stimuli had been selected based on the response of the single cells and not based on the response of human observers. This places the single cells at an advantage over the human observers, and may be partly responsible for the surprising cellular performance. The finding, that the stimuli that are hard to detect for single neurones are not the same as those that are hard to detect for human observers (demonstrated by a lack of correlation between their respective performance) shows that different stimuli would have been chosen based on the performance of human observers.

Indeed, selecting stimuli for a particular neurone indirectly makes this neurone an expert for these stimuli. Neurometric performance for this stimulus set will be high. If the brain has to perform a psychophysical decision for the same stimulus set, it would be well advised to use this particular neurone. Indeed, the single cell performance measured in this Chapter is shown to be roughly as good as that of human observers. Unfortunately, the brain cannot know *a priori* which neurone in the brain delivers such an accurate neural signal, and hence is forced to take into account also the activity of less accurate neurones. The total performance of the brain may then not be any better than that of the expert cell we selected, but the brain's performance is nevertheless not due to this particular neurone alone, but to the information contained in a population of more or less accurate neurones.

It would certainly be interesting to present RSVP sequences while a large number of neurones are recorded simultaneously. Under such conditions, the

information available in the population of neurones could be measured for stimuli that had not been optimised - for single cell nor for human observers.

The same problem applies to the study of Kovács and coworkers (1995). While they used the same stimuli for all single cells, and thus did not adjust their stimulus set to fit the response of each neurone, they nevertheless only recorded from neurones, which responded selectively to one of their stimuli. In addition, they selected the stimuli to be included in their stimulus set according to their prior knowledge of response selectivity in the region of IT they recorded in. Indeed, this is a general problem of single neurone neurophysiology: since few neurones can be recorded at once, if any results are to be obtained, stimuli must be adjusted to the neurones.

Fortunately, there is reason to believe that this argument does not jeopardise our conclusions. To perform an RSVP discrimination task on unselected stimuli, the monkey (or human) would require neurones that are as selective for those stimuli, as the neurones of Experiment 1 were for their stimulus set. The search set from which stimuli were selected in our experiment contained only about 60 stimuli, and was sufficient to activate 34 out of 137 recorded neurones. Statistically speaking, it is thus highly unlikely, that truly optimal stimuli were found for those 34 neurones. Logothetis and coworkers (Logothetis and Pauls, 1995) demonstrated that if monkeys are exposed repeatedly to arbitrary new 3D shapes, neurones could be found in IT, which were approximately as selective for these arbitrary shapes, as our neurones were for the stimuli we used. Hence it is reasonable to assume, that if an RSVP sequence was constructed from images that are somewhat familiar to the monkey, there should be neurones in the monkeys temporal cortex which are as selective for these stimuli as the single cells of Experiment 1 were amongst the 60 stimuli we used. Hence, our findings probably would not have been much different if arbitrary stimuli had been used in the RSVP sequences rather than stimuli specially selected for the individual cells - it would just have taken longer to find the neurones that would have been involved in their processing.

10.5. Conclusions

The neurometric analyses presented in this Chapter illustrate that if the responses of neurones were used to identify the presence of a particular stimulus in an

RSVP sequence, the accuracy of that decision could be very similar to that shown by human observers. Neurones in STSa therefore seem to be as smart as humans – as far as detecting the presence of a stimulus in an RSVP sequence is concerned. These findings are compatible with ascribing neurones in STSa a central role in the perception of the idendity of images such as faces. In addition the findings suggest that single neurones can be surprisingly accurate at signalling a particular stimulus. Whether STSa neurones are directly involved in *conscious* perception and perceptual decisions remains to be investigates in further experiments, but the present findings suggest that STSa may indeed be the basis for perceptual reports.

11. Experiment 3: The neurophysiology of persistence and masking in RSVP sequences with gaps.

11.1 Introduction

The literature on persistence and masking reviewed in Chapter 3 demonstrates that both physiologically, and psychologically, the representation of a stimulus does not end with the end of the physical stimulus. For some hundreds of milliseconds, the representation of the stimulus is kept alive: out of sight is not out of mind, at least not for a few hundred milliseconds. A new stimulus, presented while the representation of the old stimulus persists is thought to end the persistence of that representation. The physiological studies of Rolls and co-workers (Rolls et al., 1994a,b, 1999) indicates that 'interruption of persistence' can be observed at the level of single neurones in STSa and IT.

Despite the importance of the concept of persistence and iconic memory in psychology, there has so far been no investigation of how the 'veridical[16]' neural activity, lasting for as long as the physical presence of the stimulus, compares with the 'persisting' neuronal activity once the stimulus has been taken away from the screen. Inherent in the concept of iconic memory in psychology is the idea, that iconic memory is a form of memory which contains the same information as the sensory experience of the stimulus itself. This begs the question: is the neural activity caused by the stimulus the same as the neuronal activity persisting after the stimulus has disappeared? In particular, how does the time course of the response to a long stimulus compare to the response to a briefer stimulus (see Figure 11.1)?

---

[16] Veridical: Latin *veridicus* from *verus* 'real' or 'true' + *dicere* 'say' meaning 'saying the truth about reality'. 'Veridical' neural activity is thus meant to tell us something true about reality as opposed to persisting neural activity which in a way lies about reality: it 'speaks' about a stimulus which is no longer part of reality: reality at that moment is the black screen, and not the stimulus that used to be there before. Given cellular response onset latencies of approximately 110ms in STSa, responses always reflect an event that is part of the past. 'Persisting neural responses' refers to those responses occurring not only after the offset of the stimulus, but after the offset plus the response onset latency of the cell.

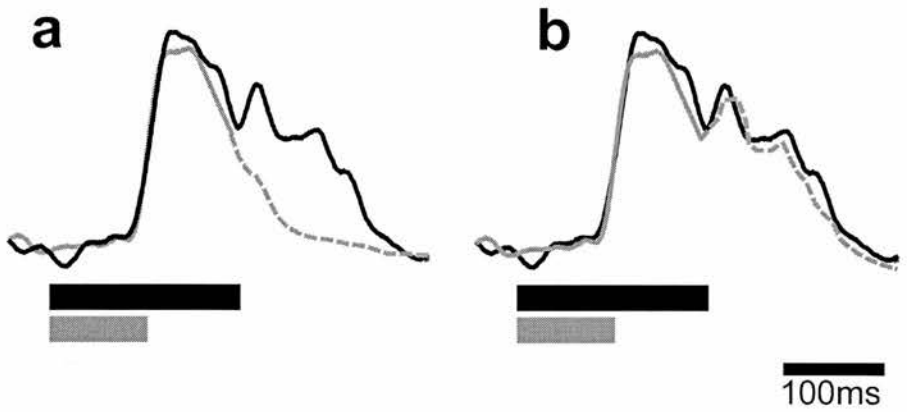


Figure 11.1: Two potential time-courses for persisting activity. The black and gray bars represent two unmasked stimuli of different duration, the lines above the bars represent potential single cell responses (sdf) to these two stimuli. (a) The persisting response (dotted grey line) to the short stimulus may decay rapidly after the end of the stimulus + response latency. Alternatively, (b) the persisting response to the brief stimulus may continue, as though the stimulus had stayed on the screen.

In the ganglion cells of the retina (Levick and Sacks, 1970, my Fig. 3.3a), the responses to a short (e.g. 16ms) stimulus are clearly different from the responses to a longer stimulus (e.g. 80ms), even if stimulus energy is kept constant. The same is true for Area 17 of the cat (Duysens et al., 1985, my Fig. 3.2c). For higher visual areas, there are presently no data indicating how the responses to brief stimulus compare to those to a longer, but otherwise identical stimulus. Rolls et al. (1994a,b, 1999), in STS and IT show that without the presentation of a masking stimulus, the activity of a short, 16ms stimulus can persist for several hundred milliseconds (see my Fig. 3.2f), but they did not measure the activity caused by a longer stimulus, and thus cannot compare the persisting activity they measured with corresponding veridical activity.

Delay activity measured in delayed matching to sample tasks is different from persistence, because it is task dependent (e.g. Fuster, 1981; Naya et al. 1996, my Fig. 3.2d,e). Delayed activity is generally smaller than the activity caused by the physical presence of the stimulus, and would thus indirectly suggest, that persisting activity too, may be smaller than veridical neural activity.

To investigate the quantitative relationship between the neural persistence to a brief stimulus and the veridical neural activity to a longer stimulus, Experiment 3 (presented in this Chapter) extended the examination of neuronal responses in STSa during RSVP sequences by introducing inter-stimulus gaps of black screen between

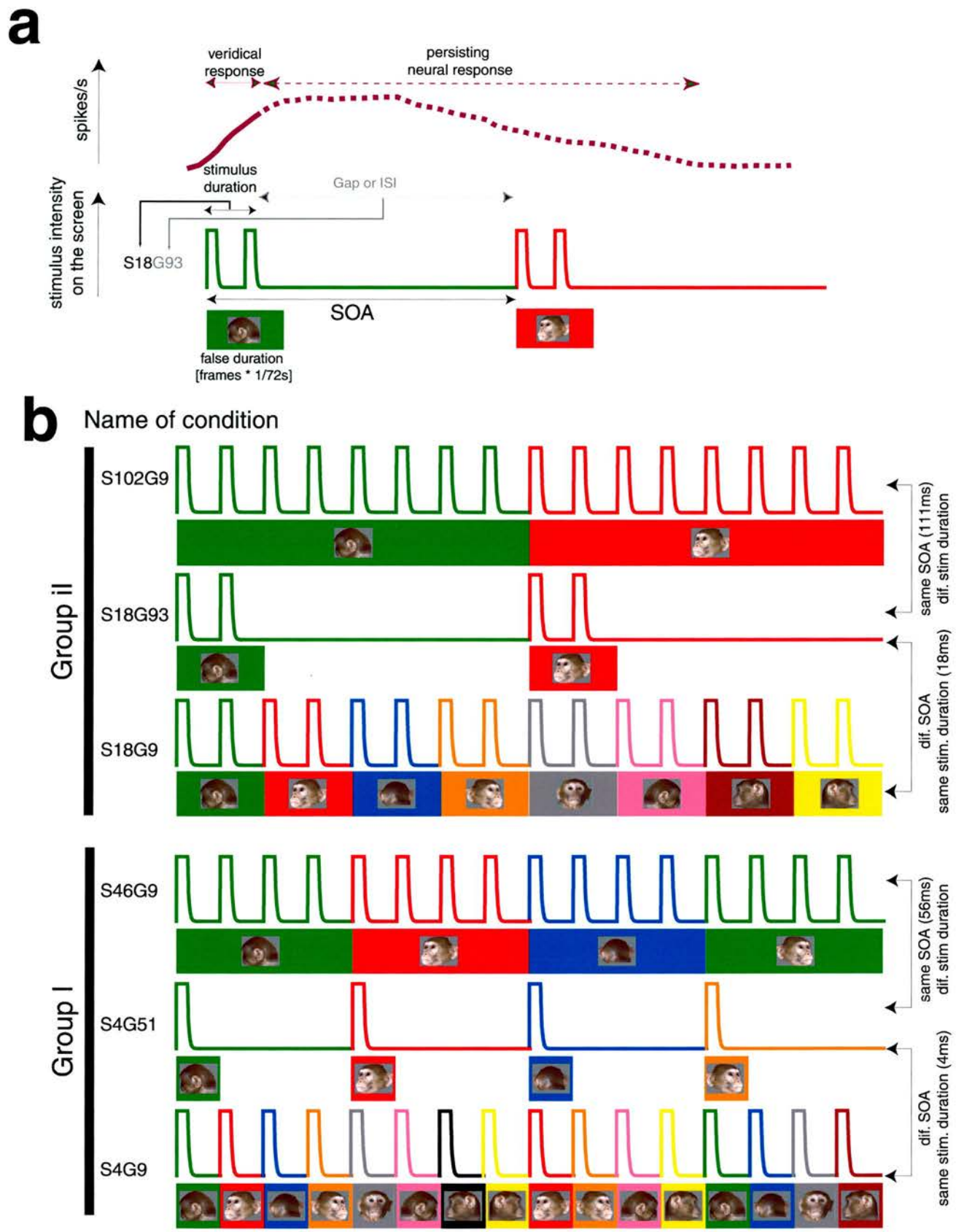


Figure 11.2: Experimental design and terminology. (a) Example of an RSVP sequence with gaps. From bottom to top: the coloured rectangles represent stimuli and white intervals represent gaps, which were black on the screen. Directly above, the actual intensity of the stimulus on the rastering monitor, incl. the 1ms phosphor decay time, as explained in Section 6.5.1. Note how this actual duration deviates from that assumed based on number of frames x 1/refresh-rate. The stimulus duration (S) measured from begining to end of the rastering process, together with the duration of gap (G) or inter-stimulus gap (ISI), all in ms, form the name of the condition (e.g. stimulus 18ms gap 93ms writes S18G93). A hypothetical neural response sdf is shown above, shifted by response latency. It can be decomposed in a veridical response (solid line), lasting for as long as the stimulus, and a persisting response (dotted), persisting beyond stimulus duration. (b) The 6 conditions used in Experiment 3. The conditions form two groups (I and II as indicated on the left). Within each group, the middle condition ('with [long] gaps') can be compared with the condition above ('long stimulus'), which has the same SOA but a much longer stimulus duration and with the condition below ('short SOA'), which has the same stimulus duration but a much shorter SOA.

each image in the sequence. Figure 11.2 illustrates the conditions used in Experiment 3.

The aim of testing the effect of short gaps (42 or 83ms long) on single cell responses was twofold.

Firstly, to examine how persisting neural activity quantitatively compares with veridical neural activity to a stimulus spanning the entire SOA.

Secondly, to clarify the interpretation of a finding of Experiment 1. Experiment 1 showed, that in RSVP sequences without gaps, the responses of a neurone to the most effective stimulus outlasted the stimulus duration by ~60ms. Given, that in Experiment 1, the onset of the new stimulus coincided with the offset of the previous stimulus, the findings of Experiment 1 can be interpreted in two ways. Neural activity in STSa may always outlast stimulus duration by ~60ms whether a masking stimulus is presented thereafter or not (fixed persistence account). Alternatively, it may be that without a masking stimulus, neural activity in STSa would outlast stimulus duration by much more than ~60ms, and that it is only the onset of the following stimulus that interrupts neural persistence, with this interruption building up over a period of ~60ms (interruption account). Introducing inter-stimulus gaps will differentiate between those two accounts: if the fixed persistence account is true, neural activity should stop ~60ms after the end of the stimulus (offset by the cells' response latency), whether the stimulus is followed by a gap or not. If the interruption account is true, neural activity should stop ~60ms after the onset of the next stimulus (again, offset by the cells' response latency), whether there was a gap or not.

11.2. Methods

General methods are as described in Chapter 6. Although the conditions with long inter-stimulus gaps truly form an experimental question separate from that of Experiment 1, the single cells data of Experiment 1 and 3 were collected together, and the conditions with 9ms gap were are the same as those reported in Chapter 7. The conditions of Group II were fully interleaved for all 34 cells tested, as described in Section 6.6.2., while the conditions of Group I were fully interleaved for 11 of the 21 cells tested in those conditions, while for technical reasons, for 10/21 cells, the

conditions of Group I were not interleaved: first condition S46G9 (i.e. stimulus duration 46 followed by the inevitable 9ms gap due to the rastering process) was collected, interleaved with the conditions of Group II. Thereafter, the conditions S4G51 and S4G9 were collected interleaved with each other.

Response analyses too were performed according to the methods described in Chapter 6, with one major exception. Initial screening of the results indicated, that response onset latency may differ between conditions with and without gaps. Response onset latency was therefore re-estimated based not on the pooled trials of all conditions except for the 14ms/image condition, but based only on conditions without gaps[17], namely the conditions formerly known as 28ms/image, 42ms/image, 56ms/image, 111ms/image and 222ms/image. This 'nogap latency' was found to be 9ms later on average than the previously used response latency (117ms vs. 108ms). N.B: this revised population latency does not change the interpretation of previous chapter.

11.3. Results

11.3.1. Visual inspection of a single neurone

Figure 11.3 illustrates the findings for a typical neurone. As SOA was reduced, both response duration and amplitude was reduced, as described in Chapter 7. Comparing the long gap and long stimulus condition at a particular SOA, i.e. comparing S102G9 (long stimulus, short gap, SOA=111ms) with S18G93 (long gap, short stimulus, SOA=111ms) or S46G9 with S4G51 (SOA=56ms), two findings are evident. First, SOA and not stimulus duration appears to determine response duration, since response duration was approximately equal for equal SOA, irrespectively of stimulus duration. Second, at equal SOA, the amplitude of the response appeared to be even larger for the *shorter* stimulus. Very clearly, conditions with long gaps show larger and longer responses than the conditions with equal stimulus duration but shorter SOA (S4G51 vs. S4G9 and S18G93 vs. S18G9).

[17] Without gaps refers to conditions without intentional gap. Even if the computer is never instructed to present a black screen frame, 9ms will elapse between the physical end of the first stimulus, and the physical start of the next stimulus.

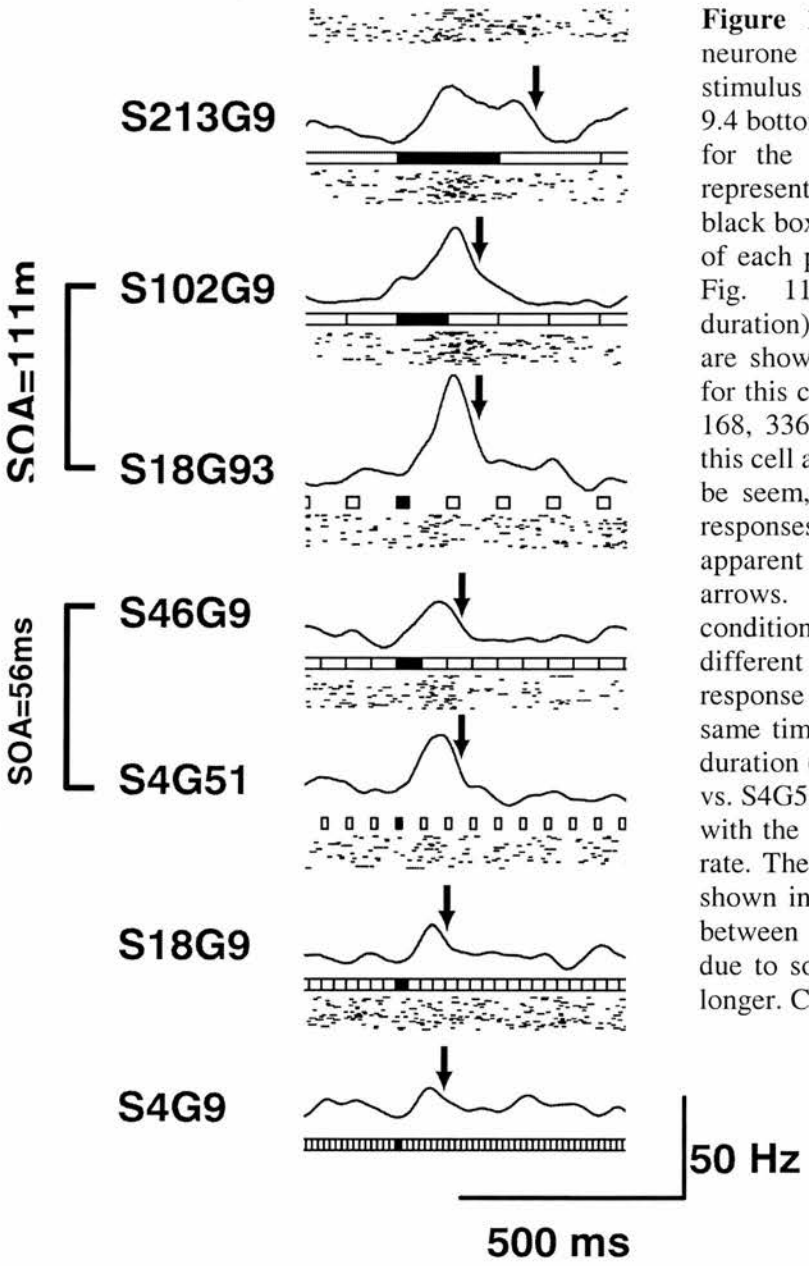


Figure 11.3: Rastergram and sdfs of a single neurone response in STSa (cell T44.5) to the best stimulus for that cell (the image of an apple, see Fig. 9.4 bottom row for an illustration of the stimuli used for the cell). The open boxes under each sdf represent randomly chosen stimuli, while the solid black box represents the stimulus of alignment. Left of each plot, the name the condition as defined in Fig. 11.2a (i.e. S_stimulus duration_G_gap duration). For comparability only the first 21 trials are shown in each rastergram. Total trial numbers for this cell from top to bottom: 21, 42, 42, 84, 84, 168, 336. All trials are considered in the sdf. For this cell all conditions were fully interleaved. As can be seem, from top to bottom, the duration of the responses decreases with decreasing SOA. The apparent end of the responses are marked by a black arrows. Comparing S102G9 with S18G93, two conditions with equal SOA (111ms) but very different stimulus duration, the offset of the response (arrow) occurred at approximately the same time, despite the large difference in stimulus duration (102 vs 18ms). The same is true for S46G9 vs. S4G51. Surprisingly, in both cases, the condition with the shorter stimulus had the larger peak firing rate. The response in the S132G9 condition is also shown in this figure to illustrate that the similarity between the S18G93 and S102G9 condition is not due to some ceiling effect: responses can still last longer. Conditions with equal SOA are marked.

11.3.2. Population analysis.

Of the 137 neurones recorded, 34 were fully tested in the 3 conditions of Group II. Only 21 of these 34 neurones were fully tested in all conditions of Group I. Since comparisons will be made between the conditions of each group, but not between conditions belonging to different groups, Group I and II will be considered separately, with Group II containing data from 34 neurones, and Group I containing data from the 21 neurones tested in all 3 conditions of Group I.

For each neurone, the 'nogap latency' was calculated. Figure 11.4 illustrates the normalised, latency aligned population sdfs in RSVP sequences with and without substantial gaps. The same general pattern apparent in the single cell of section 11.3.1. is also apparent at the population level: responses are determined by SOA and not stimulus duration; and at SOA=111ms, the S18G93 conditions causes a peak firing than the conditions S102G9.

11.3.2.1. Response duration

As in Chapter 7, a running, floating ANOVA was used to compare the population responses to the best and rest stimuli as a function of time. The right columns of Fig 11.5a,b illustrate the results of the floating ANOVA plotted on a logarithmic probability plot: the higher the curve, the more the response discriminates between the stimuli, or more precisely, the higher the curve, the less likely the responses to best and rest are drawn from the same statistical population. Table 11.1 indicates the start and end of the period of discrimination so measured.

condition	SOA	Stimulus discrimination relative to stimulus onset			Relative to latency		Discr. duration	Discr. duration-SOA	n
		peak	start	end	start	end			
S213G9	222	174	109	364	-8	247	255	33	34
S102G9	111	182	108	250	-9	133	142	31	34
S18G93	111	136	93	228	-24	111	135	24	34
S4G51	56	125	106	190	-11	73	84	28	21
S46G9	56	130	113	212	-4	95	99	43	21
S18G9	28	123	114	182	-3	65	68	40	34
S4G9	14	167	127	189	10	72	62	48	21
							average	35.3	

Table 11.1. For each condition, SOA is indicated together with information regarding the time window in which the responses to the best stimulus differ significantly from the responses to the rest stimuli, based on the floating ANOVAs illustrated in Figs 11.5a,b. All times are indicated in ms.

Based on these windows of discrimination (defined in columns 6 and 7 of Table 11.1), a window for response analysis can be defined for each Group, as lasting from the earliest start of discrimination to the latest start of discrimination for that Group, namely –11 to 95ms for Group I and –24 to 133ms for Group II, relative to nogap response onset latency.

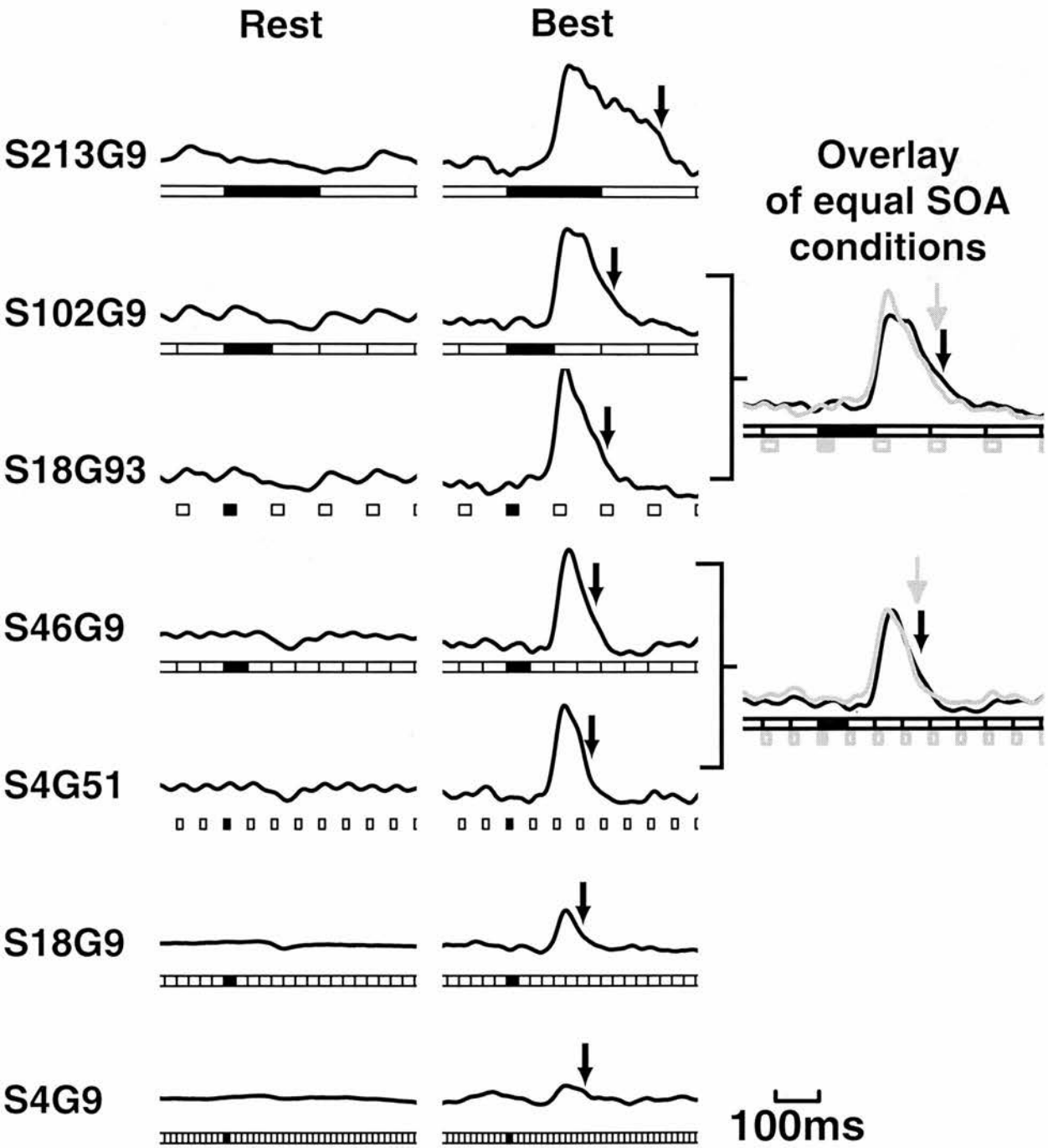


Figure 11.4: Normalised, latency aligned population spike density functions for rest and best stimuli at 7 presentation rates. The black and grey arrows indicate the end of the significant discrimination between best and rest stimuli as assessed by a floating ANOVA (illustrated in Figure 11.5). Right column: overlay of the conditions with equal SOA but different gap durations (black: 9ms; grey: 51 or 93ms). Notice how response duration decreases from top to bottom as SOA decreases, and how similar the responses are at equal SOA despite large changes in stimulus durations. Stimulus timing is indicated following the same conventions as previous illustrations.

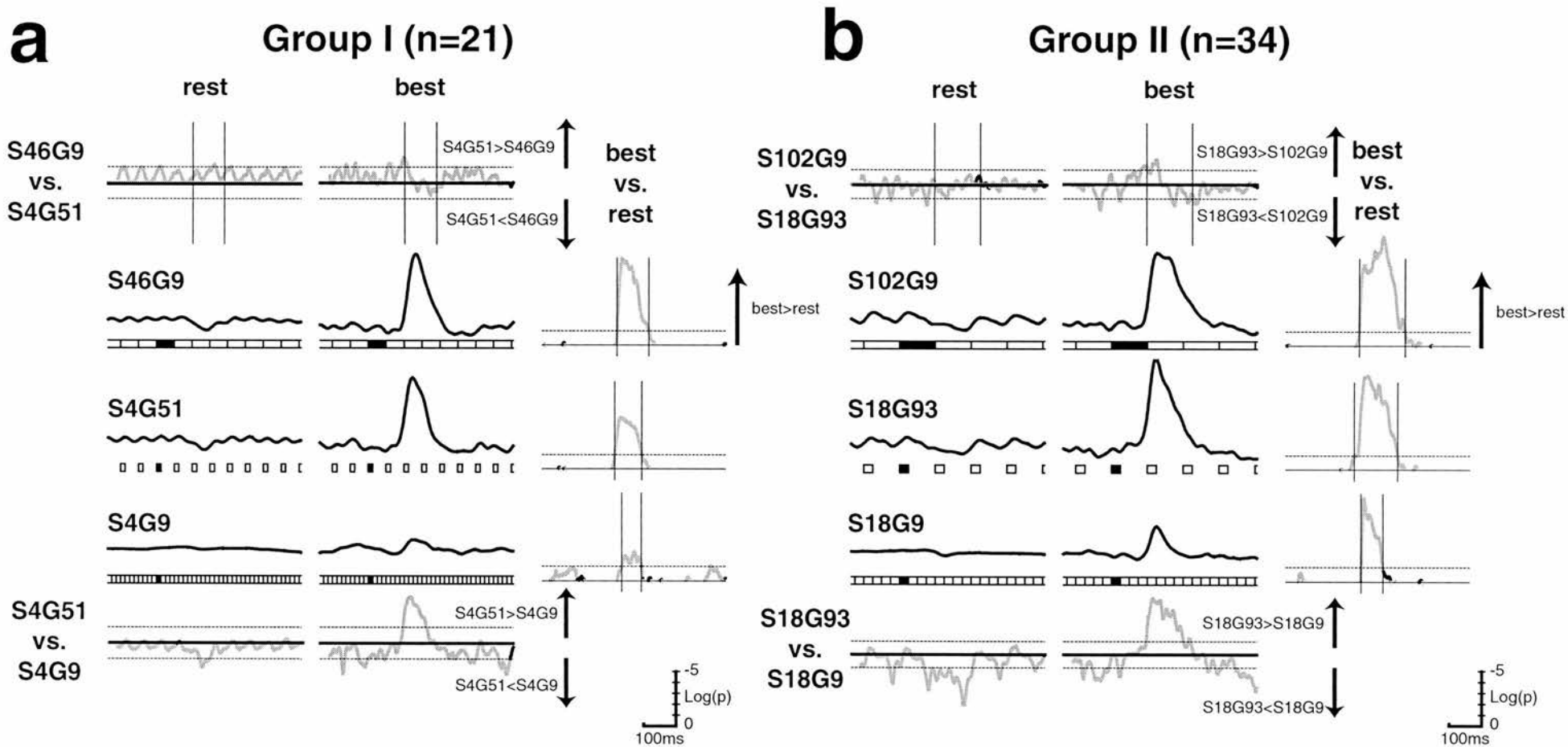


Figure 11.5: (a) The black curves represent normalised, latency aligned population sdf to the rest (left) and best (middle colum) stimuli in the 3 conditions of Group I. The grey curves on top and bottom of these sdfs represent -Log(p) of a floating ANOVA between the S4G51 condition (middle sdf) and the S46G9 (top sdf) or S4G9 (bottom sdf), for the rest stimuli (left) and best stimulus (middle column), and thus indicate how significantly the S4G51 condition differs from the two other conditions as a function of time relative to stimulus onset. The dotted horizontal lines represent the p=0.05, the black solid line the p=0 level. The direction of the deflection indicates which of the 2 conditions had larger firing rates as indicated next to the arrows. Comparing S4G9 and S4G51 in this manner confirms how much larger the responses are in the S4G51 condition. Comparing S4G51 with S46G9, shows that despite the large difference in stimulus duration, the responses are very similar at equal SOA. The vertical lines in this plot represent the beging and end of the significant best vs. rest discrimination for the S46G9 condition, and help to see how in the earliest part of the responses, the condition with the longer gap has larger responses, the opposite being true towards the end of the response. The right column contains floating ANOVAs comparing rest vs. best responses. The vertical lines represent the begining and end of the period during which discrimination is above p=0.05 (as indicated by the dotted horizontal line). Only deflections in the direction best>rest are shown in those graphs. As can be seen, discrimination starts somewhat earlier, and finishes somewhat earlier in the middle condition with the longer gaps, compared to the condition with the longer stimulus. The same is shown in (b) for Group II. The scale on the bottom right applies to the x-axes of all plots, and to the y-axes of all floating ANOVA plots (grey). (All ANOVAs are are two tailed and not boneferroni corrected).

When single cells are aligned based on the nogap latency, the floating ANOVAs reveal that responses discriminate between stimuli for on average 35 ms longer than the SOA. This is somewhat shorter than the ~60ms indicated in Chapter 7. The earliest onset of discrimination are found for the two conditions with longer gaps.

11.3.2.2. In the population, SOA and not stimulus duration determine the responses.

Both in Figure 11.4 and Figure 11.5, it appears that conditions with equal SOA but different stimulus duration produce roughly equal responses, while conditions with equal stimulus duration but different SOA produce very different responses. This is particularly evident when comparing the top and bottom probability curves of Fig. 11.5. These curves represent the probability of a floating ANOVA comparing the condition with long gaps with one of the two other conditions of each group. These probability curves are analogous to an algebraic difference of responses in that they represent the direction and amplitude of a difference, but are more sophisticated, in that the amplitude takes the reliability of the difference in the population into account. Clearly, the conditions with longer gaps resembles the condition with equal SOA but longer stimuli, and not the conditions with equal stimulus duration but shorter SOA. To test this similarity also illustrated in the overlay plots of the right column of Figure 11.4, ANOVAs were performed on the population during the window of response analysis defined in 11.3.2.1.

The surface area under the sdfs to the best stimulus for each cell was integrated during that response analysis window, an provided a single entry for each cell for each condition. These values could then be compared using matched pair ANOVAs. The results indicated in Table 11.2. fully confirm the supposed effects. If SOA is kept constant, shortening stimulus duration from 102 to 18ms (Group II) or from 46 to 4ms (Group I) did not effect the integrated responses. On the other hand, keeping the stimulus duration constant at 4 (Group I) or 18ms (Group II), and increasing the gap duration from 9ms to 51ms (Group I) or 93ms (Group II) did strongly and significantly increase the response. Hence, at the population level, the time of occurrence of the next stimulus (i.e. SOA) seems to be the critical variable determining response amplitude, while stimulus duration, within the range tested (11.5-fold change in Group I, and 5.6-fold change in Group II), had no significant effect. In addition, the amplitude of a persisting response seems overall equal to that

of a veridical response, indicating how strong neural persistence is in STSa cells, over intervals as long as 93ms.

	Conditions	What the conditions have in common	Window of analysis	F	p
Group I	S46G9 vs. S4G51	Equal SOA (56ms)	–11 to 95ms	F(1,20)=0.09	p>0.76
	S4G51 vs. S4G9	Equal stim. duration	–11 to 95ms	F(1,20)=15	p<0.001
Group II	S102G9 vs. S18G93	Equal SOA (111ms)	–24 to 133ms	F(1,33)=0.16	p>0.69
	S18G93 vs. S18G9	Equal stim. duration	–24 to 133ms	F(1,33)=51	$p<10^{-7}$

Table 11.2: Results of population comparison between pairs of conditions in each group. Conditions were compared, in a within subject, matched pair ANOVA between the surface under the sdf to the best stimulus in the two conditions, integrated in the window of analysis indicated relative to each cell's no-gap response latency.

To test whether the similarity of responses at equal SOA was also present in individual neurone responses, a cell-by-cell analysis was performed. For each cell, a spike-count distribution was computed for the best stimulus in each of the 3 conditions of each group. This was done by counting, in each uncontaminated (see section 6.7.4.) best stimulus trial, the spikes occurring in the window indicated in Table 11.2 relative to the individual cell's response latency. For each cell, a t-test (between subject design) could then be performed at alpha=0.05 (bonferroni corrected), comparing the trial-by-trial response in the different conditions, to decide whether that individual cell showed the same or different responses in two conditions. The results confirmed the results observed in Table 11.2 for the population. In Group II, comparing S102G9 vs. S18G93, not a single (0/33) neurone showed a significant difference between the two conditions; comparing S18G93 vs. S18G9, the majority of neurones (18/33) cells showed a significant difference, with the S18G93 showing larger spike counts. In Group I, comparing S46G9 vs. S4G51, again, a minority (5/21) of neurones showed a significant difference, with 2 responding more in the S4G51 condition, and 3 responding more in the S46G9 conditions; comparing S4G51 vs. S4G9, the majority of cells (13/21) responded more in the S4G51 condition, and none responded more in the S4G9 condition. These findings demonstrate that the similarity between the responses at equal SOA observed in the population can also be observed in a majority of single neurone.

11.3.2.3. Condition with longer gaps start responding ~10ms earlier.

While the overall magnitude of responses do not differ between the conditions with short and long stimuli, given that the same SOA is used, small differences in the timing of the responses become apparent under closer scrutiny. Inspection of Fig. 11.5a,b (top row) revealed that at the beginning of the response neurones respond somewhat more in conditions with longer gaps compared to conditions of equal SOA but shorter gaps. The reverse is true for the end of the response. Figure 11.6 illustrates this difference by magnifying the overlay plots of Figure 11.5 to emphasise how neurones appear to start responding about 10ms earlier in RSVP sequences with longer gaps compared to sequences with 9ms gaps.

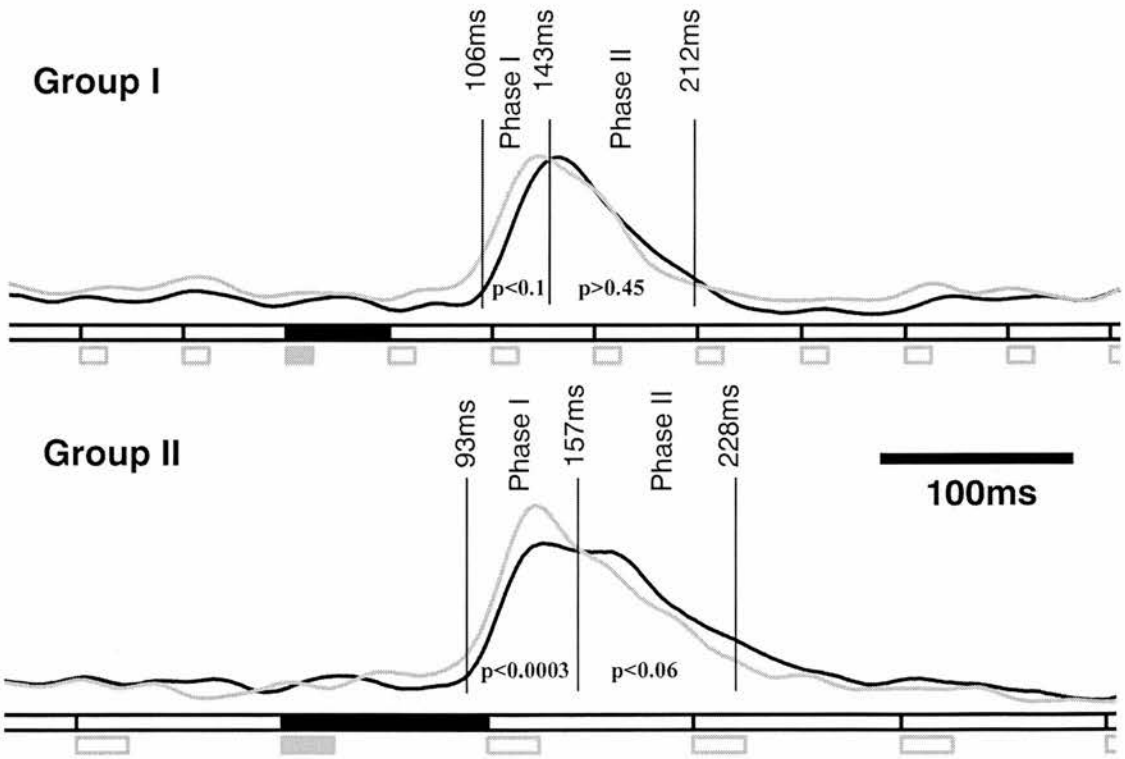


Figure 11.6: Time course of the latency aligned normalised population sdfs in RSVP presentation conditions with equal SOA, but different stimulus durations. The shorter stimulus in each group is represented in grey, the longer stimulus in black. The responses can be split in 2 phases. Phase I starts with the onset of stimulus discrimination (best vs. rest) of the short stimulus, and ends with the crossover of the average responses to the two conditions. Phase II starts with the crossover, and lasts until the end of stimulus discrimination (best vs. rest) in the condition with the long stimulus. During Phase I, the response to the shorter stimulus is larger than that to the longer stimulus while during Phase II the reverse is appears true. The p values represent the probability of the responses being equal during each phase, based on an ANOVA (see text). All timings are indicates relative to stimulus onset, and can be

expressed relative to response onset latency by subtracting the average response latency (117ms).

To test this effect, as in the preceding section, ANOVAs were performed comparing conditions based on the surface area under each neurone's sdf to the best stimulus, with each neurone providing one entry in each condition of the repeated measurement ANOVA. In Phase I (timing as indicated in Figure 11.6), the difference between the conditions S102G9 vs. S18G93 of Group II is significant ($F(1,33)=16.8$, $p<0.0003$), and the difference between the conditions S46G9 vs. S4G51 showed a trend in the same direction ($F(1,21)=2.9$, $p<0.1$). For Phase II, in Group II S102G9 produced larger responses than S18G93, with the difference almost being significant ($F(1,33)=3.8$, $p<0.06$), while in Group I, the two conditions showed almost equivalent responses in Phase II ($F(1,21)=0.56$, $p>0.45$). Overall the differences in timing are modest, except for Phase I in Group II, with a robust advantage for the condition with longer gaps.

The lack of significant differences between the conditions in Phase II indicates that the amplitude of a veridical neural response (black) is no larger than the amplitude of a persisting response for the gap duration tested (up to 93ms gap).

The larger responses in Phase I for the conditions with longer gaps go hand in hand with an earlier onset of stimulus discrimination for the conditions with longer gaps. In Group I, the floating ANOVA detected a start of best vs. rest differentiation at 113ms post stimulus onset for the condition S46G9 (Fig. 11.5b). If the responses in the condition S4G51 truly start about 10ms earlier than in condition S46G9, an ANOVA comparing the surface under the sdf for best vs. rest stimuli in the 10ms before 113ms should be significant for the S4G51 but not the S46G9 condition (each cell providing one entry). This was indeed the case ($F(1,20)=6$, $p<0.03$ for S4G51 and $F(1,20)=0.07$, $p>0.7$ for S46G9). In Group II, discrimination started at 108ms for the condition with the long stimulus (S102G9), and again, an ANOVA comparing best and rest in the 10ms before 108ms was significant in the S18G93 ($F(1,33)=5.3$, $p<0.03$) but not for S102G9 condition ($F(1,33)=0.12$, $p>0.7$), indicating that stimulus discrimination indeed starts ~10ms earlier in conditions with substantial gaps.

At the level of single cells, the earlier start of the response in conditions with longer gaps is difficult to detect. If spikes are counted in the window corresponding to Phase I, for Group II, 4 out of 33 cells show a significant difference between the two

conditions at $p<0.05$ (one tailed t-test), all of which show larger responses in the condition with longer gaps, but if a bonferroni correction of the probabilities (i.e. alpha=0.05/33=0.0015) is applied, none of the cells significantly discriminate between the conditions. For Group I, 10 cells show a significant difference, with 7/21 responding more to the condition with longer gaps at $p<0.05$ (uncorrected). If the bonferroni correction is applied, 3 cells continue to prefer the conditions with longer gaps in Group I.

11.3.2.4: Time of onset of following stimulus determines the end of the response.

As indicated by the close match of conditions with Equal SOA but different stimulus duration, it appears that the onset of the following stimulus and not the end of the stimulus of alignment determines the shape of the responses. To investigate this further, responses were aligned either on stimulus offset (i.e. the end of the stimulus of alignment) or on the onset of the next stimulus (See Figure 11.7). Clearly, when responses are aligned on stimulus offset (Fig. 11.7a), all conditions with 9ms gaps start to decrease at stimulus offset + latency, and reach baseline approximately 60ms later (grey in the figure), but conditions with longer gaps start decaying much later, and are clearly scattered, illustrating the fact that stimulus offset is not the determining variable. If instead, responses are aligned on the onset of the next stimulus (occurring 9ms later as stimulus offset for conditions with 9ms gaps), then all responses, irrespectively of the duration of their gap start to decrease at next stimulus onset time + latency, and reach baseline ~60ms later, indicating that the onset of the next stimulus is indeed the determining variable.

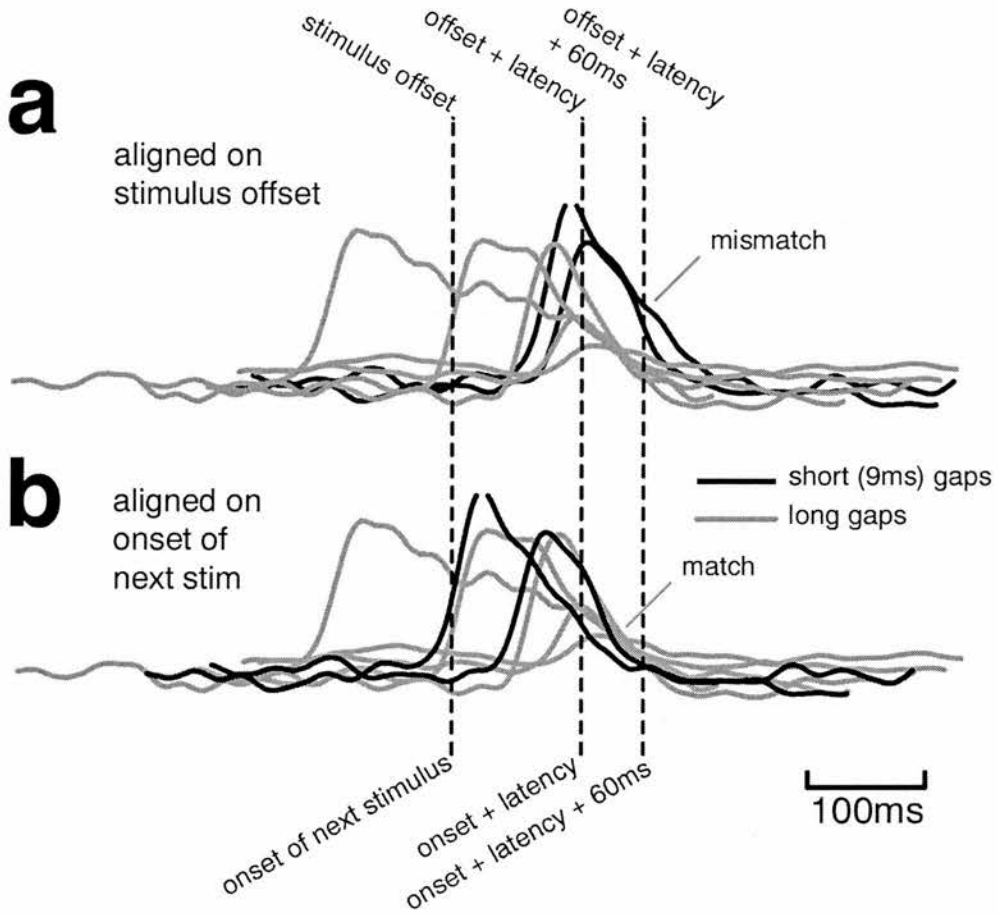


Figure 11.7: Latency aligned, normalised spike density functions for conditions with short gaps (grey, S213G9, S102G9, S46G9, S18G9, S4G9) and longer gaps (black S18G93 and S4G51) aligned on the offset of the stimulus of alignment (a) or the onset of the following stimulus (b). The event of alignment is marked by the left vertical dotted line. (a) The responses in conditions with longer gaps (51 or 93ms) clearly end later than conditions with shorter gaps (9ms) if responses are aligned on stimulus offset. In (b), all responses, including those in conditions with longer gaps start decreasing sharply at the next stimulus onset time + latency, and reach baseline ~60ms later. The match of all response-termination in (b) but not (a) indicates that the onset of the next stimulus and not the offset of the stimulus of alignment causes responses to end.

11.3.3. Information theoretical analysis

Using the methods described in Chapter 8, the single cell recordings were analysed in terms of mutual information contained between spike counts and responses ($I(S,R)$). To reduce the impact of the limited sampling bias, for each cell and group, the smallest of the 3 available trial numbers (one for each condition) was determined, and the same number of trials were selected randomly for the two other conditions of the same group. Hence trials numbers are different from cell to cell, but for each cell, trial number were identical for the three conditions of each group.

Figure 11.8a illustrates how *I(S,R)* contained in 50ms spike counts depend on the time the spike counts are taken relative to the nogap latency. In both cases, *I(S,R)* starts to rise slightly earlier in the condition with longer gaps. In Group I, the condition with shorter gaps has the slightly larger peak *I(S,R)*, while in Group II, the condition with the longer gaps has the higher peak *I(S,R)*. The peak I(S,R) was achieved in the 10-60ms window (relative to nogap response onset latency). Using that window, the two equal SOA conditions were compared in each group using a 2 condition repeated measurement ANOVA (). In both cases there was no significant difference (Group I, n=22, $F(1,21)=0.8$, $p>0.3$; $F(1,33)=1$, $p>0.3$). Results may have been more significant, if a shorter window could have been used to calculate information.

To compare the I(S,R) contained in the entire responses, *I(S,R)* was calculated for window of analysis used in Table 11.2. The results are illustrated in Figure 11.8b. For each Group, a 3 condition, repeated measurement ANOVA was performed with 22 cell entries for Group I and 34 cell entries for Group II. In both cases, the main effect of condition was significant (Group I: $F(2,42)=29$, $p<0.0001$; Group II: $F(2,66)$, $p<0.0001$). In both cases, Newman-Keuls post-hoc analysis revealed that the conditions with the shorter SOA had significantly less *I(S,R)* than the two conditions with longer SOA ($p<0.001$), but the two conditions with equal SOA did not differ significantly with respect to *I(S,R)* ($p>0.5$). Again, these findings fully confirm the impression gained based on the population response magnitudes, and demonstrates, that persisting neural responses deliver as much information about a stimulus than veridical responses – for gaps of up to 93ms.

11.4. Discussion

Single cell responses in STSa were measured while two monkeys fixated RSVP sequences with short (9ms) and longer (51 and 93ms) inter-stimulus gaps. The aim of the experiment was to investigate for the first time in the higher visual cortex, how the persisting response to a target stimulus occurring after the stimulus has ended, but before another stimulus starts, compares quantitatively with the veridical response to a stimulus covering the entire period of the same SOA (save the inevitable 9ms of gap on a rastering screen). The aim was also to test whether the end of one

stimulus, or the beginning of the next stimulus determined the duration of neuronal persistence in RSVP sequences.

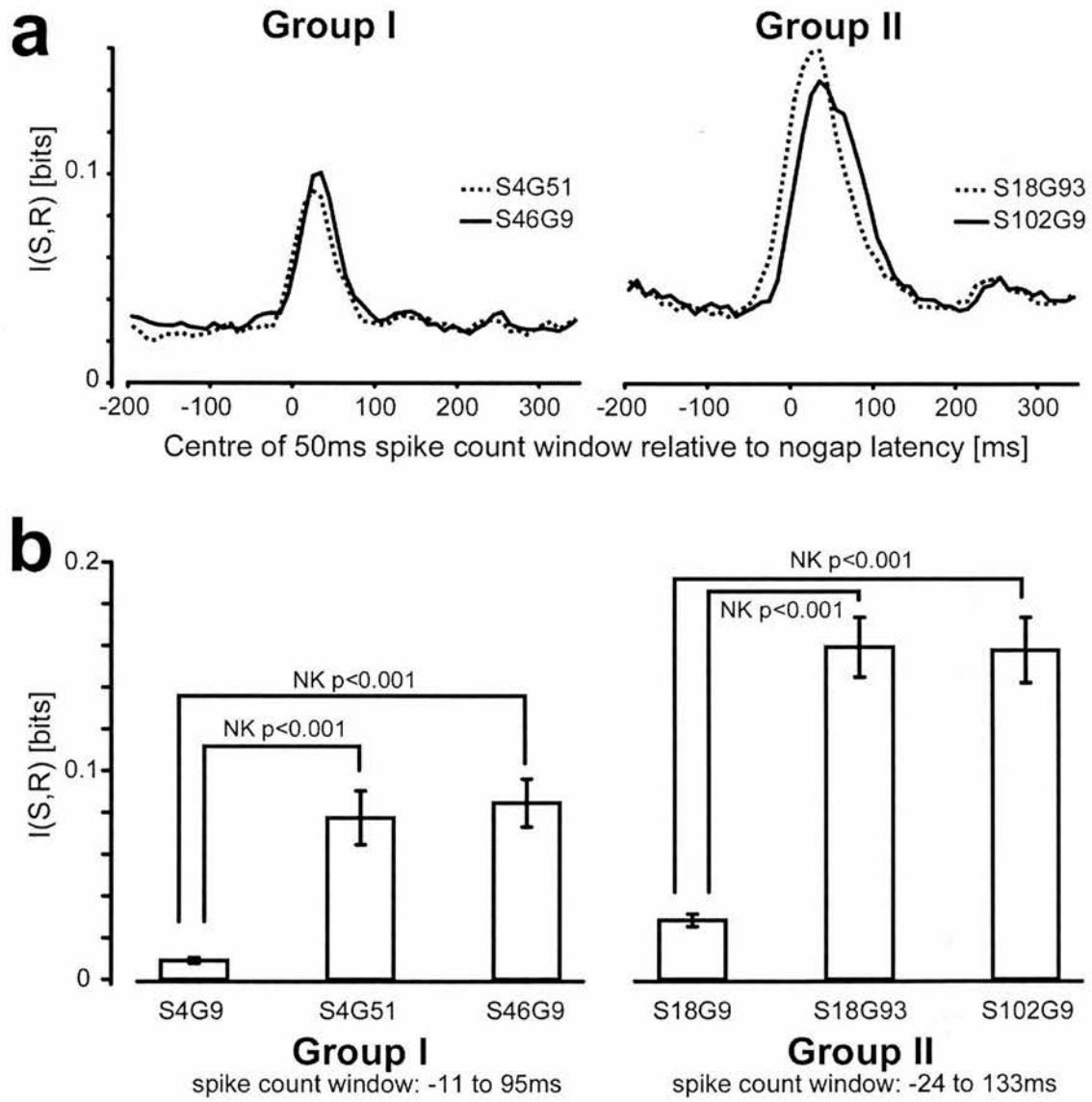


Figure 11.8: Mutual information between spike count and stimuli (*I(S,R)*). (a) *I(S,R)* contained in 50ms spike counts as a function of the centre of the spike count window for the two conditions with equal SOA of each Group. (b) *I(S,R)* contained in the entire window of analysis of Table 11.2 (as indicated below the graphs) as a function of condition. For both groups, ANOVAs indicate a main effect of condition (see text), and Newman Keuls post-hoc analyses (NK) demonstrate that the two conditions with longer SOA do not differ from each other ($p>0.5$), but that they do differ from the condition with shorter SOA ($p<0.001$).

The experiment yielded three main findings. First, single cell responses to the best stimulus in RSVP sequences are terminated ~60ms after the onset of the next stimulus, and not ~60ms after the end of the best stimulus. Second, neuronal response persistence is very strong in STSa, as strong as if the stimulus itself had stayed on the screen: at a given SOA, the persisting responses to a short stimulus followed by a gap did not differ significantly from the responses to a stimulus lasting for the entire inter-stimulus gap (save 9ms) with respect to response magnitude and mutual information between spike count and stimuli. Third, introducing longer inter-stimulus gaps enhances the early components of the responses to the best stimulus, and responses start stimulus discrimination ~10ms earlier in RSVP sequences with longer gaps compared to sequences with 9ms gaps. These points will now be discussed in more detail.

11.4.1. Persistence is terminated by the onset of the next stimulus

In accordance with the interruption theory of masking, RSVP sequences presenting a short stimulus will cause a long lasting response, which will be terminated not by the offset of the stimulus that caused the response, but by the onset of the next stimulus. This is particularly clear when one aligns the single cell responses to the end of the best stimulus. If the end of the stimulus caused the responses to decay within ~60ms (plus response latency), the responses in all conditions should decay together. This was not the case: the longer the gap, the later responses decayed relative to stimulus end. On the other hand, aligning responses to the onset of the next stimulus did cause all responses (irrespectively of gap duration) to decay within ~60ms of the onset of the next stimulus (plus response latency).

Based on estimates of response onset latency using nogap trials only, response duration outlasted SOA by 35 and not 60ms as in Chapter 7. It is likely that this difference is due to differences in the quality of latency alignment. Examining column 6 of Table 11.1 indicates that if latencies are measured based on the no-gap conditions alone, the response onset latency is somewhat misestimated: the population start discriminating between stimuli before the response onset latency! Aligning responses on a latency detected with less accuracy will add noise to the beginning and the end of the population response. This noise will reduce the stimulus discrimination in the beginning and end of the response, and thus artificially shorten the duration during

which the floating ANOVA can detect stimulus discriminating. Aligning responses on the latency detected using all conditions, including conditions with gaps, as in Chapter 7, yielded a more accurate latency alignment, as indicated by the fact that stimulus discrimination occurred on average at 108.3 ms in Chapter 7, exactly at the detected response onset. The 60ms persistence estimate of Chapter 7 should thus be more accurate than the 35ms estimate of this chapter. Indeed, Figure 11.3. illustrates how even for a single cell, response duration outlasts SOA and Figure 11.7 illustrates the decay in the population of the persisting responses over a period of ~60ms beyond the onset of the next stimulus + SOA.

The finding, that this progressive decay over ~60ms applies not only to conditions in which the stimulus ended just before the onset of the next stimulus but also to conditions in which the stimulus itself had ended 93ms earlier confirms the idea, that for STSa neurones, persisting responses are no different from veridical responses (for persistence intervals of up to 93ms at least).

Overall these findings strongly support the idea that masking is due to the competition between the representations of sequentially presented stimuli. Masking really is 'temporal rivalry'.

11.4.2. STSa does not differentiate between persisting and veridical responses

One of the main motivations of the present study was to compare quantitatively persisting responses with veridical responses: how does activity associated with the iconic memory of a stimulus compare with the activity to a continuing stimulus (Fig. 11.1)? The answer is that as far as STSa cells are concerned, and for inter-stimulus gaps of up to 93ms, there is no significant difference between veridical and persisting neural activity. This holds true for if the neurones are considered as a population, but also for the majority of single neurones, and holds true wether the response magnitude or the mutual information between responses and stimuli is considered.

These findings are surprising when compared with findings of Levick and Sacks (1970, my Fig. 3.3a) and Duysens et al. (1985 my Fig. 3.2c). Both studies show, that up to area 17, neural persistence after a short unmasked stimulus differs from the response to a continuing stimulus. This difference could be due to species differences: the visual system of cats may be more specialised in the processing of motion, where neural persistence may not be adaptive. Monkeys, having gone though

an evolutionary phase of frugivority, may have optimised their ventral visual system for fine, static shape processing, where persistence may be more adaptive. Measurement of persistence in earlier visual areas of primates would test this hypothesis.

It may also be, that visible persistence is more pronounced, not only in *species* that specialise in the processing of static form, but also within one species, in *brain areas* more specialised in shape processing. Neurones in the temporal cortex specialising in shape processing may show more visible persistence than single neurones in earlier areas. This would make sense, considering that earlier areas have to feed not only the ventral but also the dorsal visual stream. For the processing of motion, it is easily conceivable that visible persistence would lead to blurring and would be detrimental. These early visual areas may thus have to be more veridical in their representation of the physical duration of a stimulus. The strong visible persistence observed in STSa may thus be a luxury of the parallel, distributed visual system: while other areas 'take care' of extracting the precise timing of a stimulus, STSa can have the luxury of not caring about the physical termination of a stimulus, but rather it may continue processing a stimulus until it is interrupted by input regarding a new complex stimulus. This strategy would allow extracting as much information about the shape and identity of a briefly presented stimulus as possible, without loosing temporal information about the stimulus, since temporal information would be processed by different cortical processing areas.

It would be interesting to repeat Experiment 3 in different areas along the ventral stream (e.g. V1, V2, V4, PIT, CIT, AIT), using stimuli appropriate for driving single neurones at those levels, and see how neural persistence changes along the ventral stream or between ventral and dorsal stream.

Interestingly, the black screen during the inter-stimulus -gap is not an effective stimulus in terms of interruption, since conditions in which a brief target stimulus was followed by a gap produced longer responses than if it was followed by an other complex image. This reinforces the notion of object competition in the ventral stream: a black screen is not an effective stimulus at driving neurones higher up in the ventral stream. If the black screen is not represented as an event in the temporal cortex, it *a fortiori* cannot compete with the representation of the target stimulus, and hence cannot terminate the responses.

Finally, the findings presented here hold for gaps of up to 93ms, and will of course not hold true for indefinitely long gaps. To determine when persisting responses become significantly smaller than veridical responses, single neurones are currently being recorded in a further comparison Group: S46G9 vs. S46G176 vs. S213G9. This group contains two conditions with SOA=222ms, and it is expected, that for a gap duration of 176ms, differences will arise between veridical and persisting responses. Unfortunately, sufficient data is not yet available.

11.4.3. Forward masking and the increased responses for conditions with long gaps

One would assume that shortening stimulus duration while leaving SOA unchanged should reduce the amplitude of the responses. This is particularly true, because stimulus energy in Experiment 3, according to Bloch's law (stimulus energy=stimulus intensity x duration for durations up to ~60ms), was about 3x larger in the S102G9 compared with the S18G93 condition, and about 5x larger in the S46G9 compared with the S4G51 condition.

The findings of this experiment do not support this assumption. The overall response magnitude is the same at equal SOA irrespectively of stimulus duration. Single neurone in STSa thus seem 'not to care' about changes in stimulus energy, a finding confirming the earlier findings of Rolls and Baylis (1986) and Hietanen, et al. (1992), that responses of single cells in STSa are relatively insensitive to changes in contrast or illumination.

Indeed, not only is the overall response magnitude, measured as the integration of the response over the period of stimulus discrimination, not smaller in conditions with shorter stimulus duration at equal SOA, but the responses are even start about 10ms earlier than in conditions with 9ms gaps. In addition the early components of the response in conditions with 93ms gaps are *larger* than responses with 9ms gaps but much longer stimuli. Together this effect can be called a 'gap-advantage': single neurones in RSVP respond earlier, and more vigorously, if the senquences contain gaps, even at the expenses of stimulus duration. Quite the opposite of what would be expected based on stimulus energy: neuronal response onset latencies usually diminish with increasing stimulus contrast (e.g. Gawne et al., 1996), and it would be expected, that stimulus energy should have similar effect.

The information theoretical analysis also showed the same trend, but differences in peak information were not significant in either Group. It is possible,

that using smaller spike count windows would have allowed to demonstrate a difference in the peak information in Group II. Unfortunately, calculating information using spike count windows smaller than 50ms tends to yield very noisy results. Alternatively, it may be that significant difference in the peak response between S102G9 and S18G93 may not have resulted in a significant difference in peak information because responses may be more variable in the S18G93 condition. Figure 11.5 indicates that it was not due to the response to the rest stimulus being larger in the S18G93 condition.

Forward masking may be responsible for this gap-advantage detected based on response magnitude. Little attention has been given to forward masking in the visual system, but Macknik and Livingstone (1999) demonstrated that forward metacontrast masking reduces the transient on response of single neurones in V1, and that the offset and not the onset of the forward masking stimulus is the masking event. In addition, the seminal work of Crawford (1947) illustrates how both the on- and off-transient of a stimulus can have masking effects on a following stimulus (see my Figure 3.7). In the RSVP experiment presented here, in conditions with 9ms gap, the offset of the preceeding stimulus is only 9ms away from the onset of stimulus of alignment, and that temporal proximity enables strong forward masking. As gap duration is increased, the offset of the preceding stimulus is increasingly further away from the onset of the stimulus of alignment, and decreasingly forward masking effects would be expected. The early components of the response to the stimulus of alignment should be particularly sensitive to forward masking, and should thus increase in amplitude as the gap duration is increased. This is exactly what was measured in our experiment. Both the fact that stimulus discrimination arose 10ms earlier in conditions with longer inter-stimulus gaps and the larger peak-firing rate visible in Figure 11.3-7 for the condition with 93ms gaps can thus be accounted for by forward masking through the offset-transient of the preceding stimulus. Presenting pairs of images to monkeys while single STSa neurones are recorded and varying the ISI between the two images of each pair could test this account: in trials where the second stimulus produces a significant response, one would expect a decrease in response latency and an increase in peak firing rate as the ISI is increased.

11.5. Conclusions

The present investigation of the effect of inter-stimulus gaps in RSVP sequences on the responses of single neurones in the STSa demonstrates how powerful neuronal persistence can be in STSa neurones: the persisting response to a short stimulus during an inter-stimulus gap is indistinguishable from the veridical response to a stimulus lasting throughout the same SOA period – both in terms of response magnitude and mutual information. This makes STSa neurones unsuitable for supporting judgments regarding the exact duration of a stimulus, but makes STSa ideally suited for extracting as much information about the shape of a briefly presented stimulus as possible. If STSa neurones are indeed involved in the perception of the identity of the object they respond to, and if subjects would indeed base their decision in a psychophysical detection task on the firing of STSa neurones, as suggested in Chapter 10, then one would expect psychophysical performance at equal SOA to be equal irrespectively of stimulus duration: a hypothesis that will be tested in the next chapter.

12. Single cells and perception of identity – the effect of inter-stimulus gaps

12.1. Introduction

Chapter 11 demonstrates that the magnitude of the responses of single neurones to individual images in RSVP sequences does not depend on the duration of the stimuli given that SOA is kept constant. Presenting each stimulus for 4ms followed by a gap of 51ms for instance produced responses of the same magnitude as presenting each stimulus for 46ms followed by a 9ms gap.

This finding is intriguing, because to a human observer reducing the duration of a stimulus while keeping the SOA constant greatly changes the percept associated with the RSVP sequence: the perceived contrast of the images is reduced (Bloch's law), and a clear flicker becomes increasingly perceivable (see Movie 12.1 on the attached CD)

If STSa neurones are associated with visual perception and visual decision taking as hypothesised based on the similarity between neurometric and psychophysical performance in Chapter 10, one would expect the perceptual changes experienced as gap duration is increased to be associated with clear changes in the responses of STSa neurones. Such changes were not apparent in the single cell recordings of Chapter 11 if the cells were considered as a population, and were only rarely apparent if neurones were considered individually: not a single neurone responded significantly differently in the S18G93 and the S102G9 condition, and only 5/21 neurones responded significantly differently in the S4G51 and the S46G9 condition. Even the latter 5 neurones showed effects in opposite directions: 3 responded more in the S46G9 and 2 in the S4G51 condition. This overall lack of response differences in STSa neurones as stimulus duration is varied poses an obvious challenge to the idea of STSa neurones conveying the perception of complex stimuli such as faces. Equal STSa responses should create equal percepts, yet as stimulus duration is changed but SOA is kept constant, STSa responses remain equal, while the percept changes.

To help understand this problem, it is important to distinguish different aspects of the perception of an image. If I would present my holiday photographs to you in a slide presentation, and two days later show you the same photographs as hard copies in my dark cellar office, the photographs would physically look different to you in many ways, and yet you would resent me for boring you again with the same photographs. What changed between the slide show and the hard copies of the same photographs are what I will call 'low-level' properties: the contrast, the brightness, the size, and the presentation duration of the images. None of these low-level properties prevents you from realising that I am boring you with the same photographs. What did not change between the two presentations is what I will call the 'identity[18]' of the photographs. By that I mean the identity of the *photographs* and not the identity of the object being depicted. You may not recognise, that one of the slides was a photograph of the table mountain in Cape Town (South Africa), but you are still able to recognise that the hardcopy of this photograph is the same *photograph*. The identity of the object being depicted may remain unidentified, while the identity of the photograph can be recognised. Identity will be used here, in the meaning of identity of the photograph. Judging the identity of a photograph in a recognition test can often be achieved based on individual, distinctive features, such as a particular shadow on the ridge of the mountain, and need not be based on the knowledge of what is depicted.

It is possible, that the perception of identity and the perception of low-level properties may be mediated by different neural substrates STSa neurones could participate in perception of identity without reflecting strongly perceived low-level properties. Indeed, Zeki and Bartels (1999) propose that visual consciousness is not unified. Instead, they propose that each cortical area contributes independently to consciousness. The area MT, specialising in the processing of motion, could provide the percept of motion, while V4 specialising in the processing of colour could provide the percept of colour. Zeki and Bartels call each of these individual pieces of consciousness 'micro-consciousness'. The micro-consciousness mediated by each cortical area is thought to reflects the properties of the visual world that are explicitly represented in the neural responses at that stage of neural processing. By explicit, they mean requiring no further processing.

---

[18] Late Latin *identitas*, from Latin *idem*, 'the same'

The present Chapter will explore the hypothesis that STSa may mediate perception of identity while other cortical areas mediate the perception of 'low-level' properties such as flicker and contrast. The means to this aim will be to measure human performance in two psychophysical tasks which require different attributes of image perception to be attended to. The tasks will be performed using RSVP presentation conditions similar to those used to test single neurones in the previous chapter. Using the neurometric methods from Chapter 10, one can then identify attributes of visual perception that are associated[19] with and attributes that are dissociated[20] from neuronal responses in STSa.

12.2. Experiment 4 – The psychophysics of image identity

12.2.1. Introduction to Experiment 4

Experiment 4 of this thesis is a replication of the detection task of Chapter 9 using the presentation conditions of Chapter 11. To perform the detection task (see Figure 9.1), it is necessary to match the identity of the image, and to ignore low-level properties: the target is presented first for 300ms, outside of a sequence, and then for as short a duration as 4ms in the sequence. It is obvious, that those two occurrences will differ in low-level properties such as duration and apparent contrast. Hence this task is suited to test how the perception of identity will be affected by the long inter-stimulus gaps of conditions S4G51 and S18G93 (see Figure 11.2).

In a second step, the single cell data of Experiment 3 will be analysed using the neurometric methods used in Chapter 10. This will indicate how the performance of an ideal observer of the activity of single STSa cells is affected by variation of stimulus duration at equal SOA.

Finally, human performance will be compared with the supposed performance of the ideal observer of STSa activity to test if perception of identity can be associated[2] with STSa activity.

---

[19] By 'associated with', I mean that changes in presentation conditions affects the performance of human subjects in a task in the same way as it affects the performance of an ideal observer of the firing rate of single cells in STSa.

[20] By 'dissociated from', I mean that changes in presentation condition affect the performance of human subjects in a task in a way that differs from the way it affects the performance of the ideal observer of STSa activity.

12.2.2. Methods of Experiment 4

Stimuli and procedures: Experiment 4 is essentially a replication of the detection task of Experiment 2, and the reader should consult Chapter 9 for details of the methodology. The following differences apply:

First, different presentation conditions were used, namely those of Experiment 3, illustrated in Figure 11.2: 2 groups of 3 conditions each, with Group I composed of conditions S4G9, S4G51 and S48G9 and Group II composed of conditions S18G9, S18G93 and S102G9.

Second, of the 23 stimuli used in Experiment 2, all were tested again in Experiment 4, but only 21 were analysed. This was so because only 21 cells could be analysed fully with neurometric methods in all 6 conditions of Experiment 3. Hence, to allow a 1:1 comparison between psychophysical and neurometric data, only those stimulus sets for which full neurometric analysis exists were taken into account for the psychophysical results.

Subject: In Experiment 4, myself (CK) and DX served as observers. In Experiment 2 naïve subjects did not differ clearly from CK and DX, and hence it appeared unnecessary to recruit naïve subjects for this study. Both subjects had extensive experience with the stimuli and the psychophysical task from having completed the psychophysical task earlier in Experiment 2.

Data analysis was performed using the methods described in Chapter 9 and 10.

12.2.3. Results of Experiment 4

12.2.3.1. Psychophysical results

The behavioural performance was analysed separately for each stimulus set and presentation condition by pooling the results of the two experimental subjects and creating a signal detection table. The signal detection table was then used to compute a proportion correct decision and a mutual information value between behavioural response and stimulus presence. Results are illustrates in Figure 12.1a for proportion correct responses and 12.1b for mutual information.

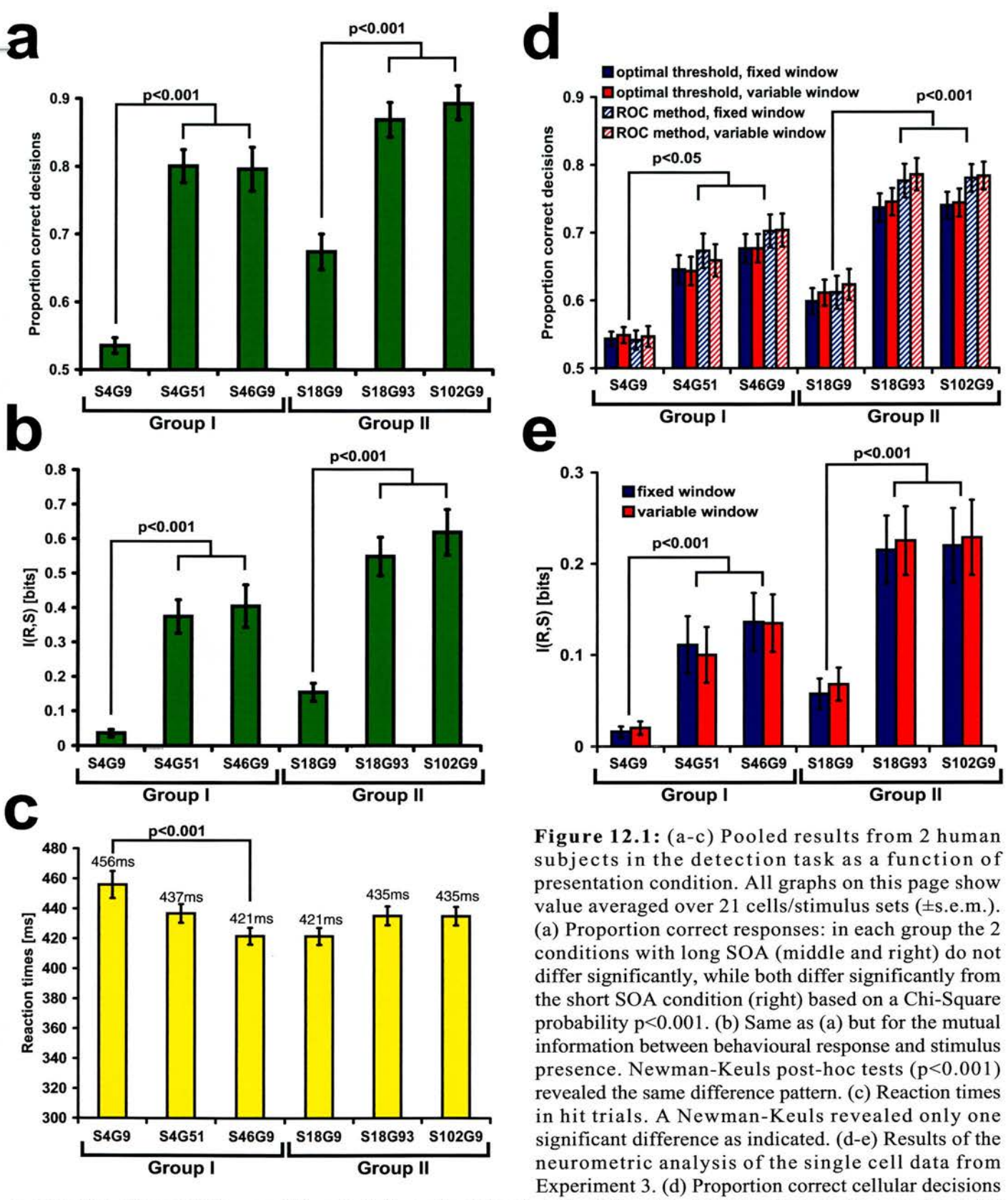


Figure 12.1: (a-c) Pooled results from 2 human subjects in the detection task as a function of presentation condition. All graphs on this page show value averaged over 21 cells/stimulus sets (±s.e.m.). (a) Proportion correct responses: in each group the 2 conditions with long SOA (middle and right) do not differ significantly, while both differ significantly from the short SOA condition (right) based on a Chi-Square probability $p<0.001$. (b) Same as (a) but for the mutual information between behavioural response and stimulus presence. Newman-Keuls post-hoc tests ($p<0.001$) revealed the same difference pattern. (c) Reaction times in hit trials. A Newman-Keuls revealed only one significant difference as indicated. (d-e) Results of the neurometric analysis of the single cell data from Experiment 3. (d) Proportion correct cellular decisions as a function of presentation condition. Both the optimal threshold (solid) and the ROC (hashed) method are shown for both the fixed (blue) and the variable (red) spike count window. The overall finding is the same in all 4 cases: the short SOA condition (left) is smaller then both the longer SOA conditions (middle and right) within each Group, while conditions with equal SOA (middle vs. right) do not differ significantly, as indicated by a Newman-Keuls post-hoc analysis ($p<0.001$). (e) Same as (d) for the mutual information between cellular response at optimal threshold and stimulus presence. Comparing (a) with (d) and (b) with (e) illustrates how human subjects and single cells in STSa demonstrate the same pattern of results: increasing SOA increases neurometric and human psychophysical performance, but modifying the duration of a stimulus at equal SOA does not significantly effect neurometric or human psychophysical performance in detecting stimulus identity.

If STSa activity is associated[2] with the perception of stimulus identity, based on the results of Chapter 11 that the responses of STSa neurones are unaffected by variations of stimulus duration at equal SOA, human performance in the detection task should be similar for conditions having equal SOA but different stimulus duration.

Figure 12.1a clearly supports this prediction: detection performance appear to be equal for presentation conditions of equal SOA but different stimulus duration. Reducing SOA on the other hand reduces performance. To test these effects statistically, proportion correct decisions were compared pair-wise using the Chi-Square non-parametric statistics: the number of correct and incorrect trials in one condition were compared with the number of correct and incorrect trials in the other condition. For instance in the S4G9 condition, the two subjects totalled 360 correct and 312 incorrect trials (pooled over 21 stimuli with 16 repetitions each x 2 subjects), while in the S4G51 conditions, they totalled 538 correct and 134 incorrect trials. The Chi-square statistic indicates that the probability of the two scores being identical is less than 0.001 (χ^2=106.32, df=1, p<0.001), confirming that the S4G9 condition was significantly different from the S4G51 conditions. On the other hand the S4G51 condition did not differ from the S46G9 condition (χ^2=0.04, df=1, p>0.83). In Group II, S18G9 differed significantly from S18G93 (χ^2=72.45, df=1, p<0.001) but S18G93 did not differ from S102G9 (χ^2=1.82, df=1, p>0.17). The psychophysical results therefore indicate that leaving stimulus duration constant but prolonging SOA facilitates the detection of a particular target image identity in RSVP, while changing the duration of a stimulus but leaving SOA constant, does not influence identity detection performance for SOAs of up to 111ms.

To investigate if the similar performance in the S4G51 and S46G9 condition could have been due to a ceiling effect, the proportion correct in those two conditions was compared with the performance correct in the S18G93 and S102G9 condition. Performances were compared pair-wise using the Chi-Square statistics, and S4G51 was shown to lead to smaller performance than both S18G93 and S102G9 (both χ^2>11, df=1, p<0.001). Similarly, the S46G9 was shown to yield lower proportion correct trials than both S18G93 and S102G9 (both χ^2>13, df=1, p<0.001). It is thus possible to achieve significantly better detection performance than in the S4G51 and

S46G9 condition and the similar performance in the two latter conditions cannot be due to a ceiling effect.

The psychophysical performance in the detection task was above chance at all conditions, even in the S4G9 condition: A Binomial distribution with p=0.5 indicates that the 360 or more correct trials out of a total of 672 trials occurs with a probability $p<0.03$. Experiment 4 therefore replicates the findings of Experiment 2 of Chapter 9, namely that human observers can detect a target image above chance in RSVP sequences presented at 14ms/image, and demonstrates that this effect can be measured even in a small sample of 2 subjects.

Since the same 23 stimulus sets had been measured using some of the same conditions in Experiment 2 and 4, a test-retest reliability was calculated using all 23 stimuli. Four conditions were common to the two experiment, each with 23 proportion correct detection scores (one per stimulus set, in this case all 23 stimuli were analysed even in Experiment 4 since no comparison with the neurometric results is necessary). Those 4x23 values were correlated between the two experiment, and the correlation was high ($r=0.83$ if all 5 subject are considered for Experiment 2, and $r=0.82$ if only the two subjects tested in both experiments are considered) indicating that the detection task yielded reliable estimates of human detection performance. If only the two subjects tested in both experiments (DX and CK) were considered, the performance, averaged over all presentation rates, was 1% better in Experiment 4 compared to Experiment 2 (73% correct vs. 72% correct), but the difference was not significant (matched pair t-test, $t(91)=0.91$, $p>0.36$) indicating that although both subjects used for Experiment 4 had participated 6 month earlier in Experiment 2, no significant learning effect was observed.

If mutual information was analysed instead of the proportion correct responses, results remained identical. Mutual information was analysed using a 2 group x 3 condition repeated measurement ANOVA, with each of the 21 stimulus sets entering one proportion correct in each case. The ANOVA yielded a significant main effect of group ($F(1,20)=49$, $p<0.001$), indicating that information was higher in Group II than in Group I, and a main effect of condition ($F(2,40)=57$, $p<0.001$). A Newman-Keuls post-hoc analysis (critical range = 0.05) indicated that the main effect of condition was due to the condition with the shorter SOA being different from both the two conditions with longer SOA (both $p<0.001$), while the two conditions with

equal SOA did not differ with respect to mutual information ($p>0.2$). The interaction between group and condition was not significant ($F(2,40)=1.33$, $p>0.27$).

To investigate the possibility of a speed accuracy trade off, reaction times for hit trials were averaged for each stimulus set at each presentation rate (Figure 12.1c). ANOVA and Newman Keuls post-doc indicated that the only significant difference was between the S4G9 and S46G9 condition (Newman-Keuls, $p<0.05$), but the difference was in the direction opposite to a speed accuracy trade off, as RT was longest in the condition with the lowest accuracy.

Also there was no significant difference between the RT for conditions with the short and long gaps at equal SOA. This is surprising given that for the single cells, conditions with longer gaps produced ~10ms earlier responses. RT on the other hand were exactly equal for the S18G93 and the S192G9 condition, and Group I reactions times, although not significantly different, were even non-significantly 16ms *longer* in the conditions with the longer gaps.

Overall, the psychophysical results of experiment 4 therefore fully confirm the hypothesis derived from the cellular responses of Experiment 3, that at equal SOA, detection performance for image identity remains equal despite changes in stimulus duration, but that performance changes if SOA is changed.

12.2.3.2. Neurometric analysis of Experiment 3

As in Chapter 9, both the optimal threshold method and the ROC method were used to calculate the proportion correct cellular decisions. Again, either the same window was used for all conditions of one group, or each condition was associated with a unique spike counting window. In the former case, spikes were counted in uncontaminated target present and target absent trials from –11ms to 95ms (Group I) or from –24 to 133ms (Group II) relative to each cell's nogap latency, and the window was called "fixed window". Using this window ensured that any difference in the proportion correct observed within a group could not be due to differences in the size of the spike-counting window. As an alternative, a 'variable window' method was used, where spikes were counted in a different window for each condition. These variable spike counting windows were equal to the periods during which the population of neurones discriminates between best and rest in each condition, as indicated in Table 11.1, relative to each neurones nogap response onset latency.

12.2.3.2.1. Proportion correct responses

Figure 12.1d illustrates the proportion correct cellular decisions obtained for 21 neurones using the two methods and the two time windows of analysis. To analyse the results a within subject, 2 method (optimal threshold vs. ROC) x 2 windows x 2 groups (Group I vs Group II) x 3 conditions (short SOA vs. long gap vs. long stimulus) ANOVA was performed. The two methods yielded very similar results, but the ROC method gave consistently larger proportion correct cellular decisions, as indicated by a significant main effect of method ($F(1,20)=50.7$, $p<0.001$). Method also interacted significantly with condition ($F(2,40)=26$, $p<0.001$). The spike-count window was of little importance, and had no significant main effect ($F(1,20)=4.08$, $p>0.05$).

As for the detection task, both group and condition had highly significant main effects (group: $F(1,20)=58$, $p<0.001$; condition: $F(1,20)=61$, $p<0.001$). The main effect of condition was analysed using a Newman-Keuls post-hoc test, which revealed that the long gap and long stimulus condition did not differ significantly from each other ($p>0.05$), but that the short SOA condition differed from both the long gap and long stimulus condition (both $p<0.001$). Again, there was no interaction between group and condition, ($F(2,40)=1.8$, $p>0.17$), indicating that the effect of condition was similar in both groups.

12.2.3.2.2. Mutual information

If the mutual information between cellular decision at optimal threshold and stimulus presence was analysed instead of proportion correct, the results were identical to those obtained using the proportion correct: there was a main effect of window ($F(1,20)=7$, $p<0.02$), with the variable window yielding higher information, a main effect of group ($F(1,20)=19$, $p<0.001$) with Group II having higher information, and a main effect of condition ($F(2,40)=18$, $p<0.0001$). A Newman-Keuls indicated that again, the main effect of condition was due to the condition with short SOA having lower information values than the two conditions with longer SOA (both $p<0.001$), but the two conditions with equal SOA did not differ significantly ($p>0.5$).

Overall, these neurometric analyses indicate that an ideal observer's decision about the identity of a particular stimulus in a sequence based on the firing of single neurones in the STSa is not affected by stimulus duration as long as SOA is kept

constant. Reducing the SOA on the other hand significantly reduces the performance that can be achieved based on the responses of STSa neurones. These findings confirm what was apparent in the magnitude of the responses in STSa.

12.2.3.3. Comparing Neurometric and Psychometric results

When comparing the psychometric and the neurometric results presented above, it appears that the performance of single cells is slightly less accurate than that of human subject in the detection task. This finding confirms the results of Chapter 10, and is partially attributable to the beneficial effect of selective attention in the detection task, as discussed in Chapter 10.

If the proportions correct cellular decisions are compared with the proportions correct decisions in the psychophysical task, using a 2 PN (psychometrics vs. neurometrics) x 2 group x 3 condition repeated measurement ANOVA, performed separately for the 2 methods (optimal threshold vs. ROC) and the 2 windows (fixed vs variable), the results are always the same: there is a main effect of PN (all $F(1,20)>12$, all $p<0.002$), with the human performance being superior to the cellular performance, and there is always a significant PN x condition interactions (All $F(2,40)>8$, all $p<0.001$), indicating that the difference between the single cells and the human observers is larger for longer SOA – just as in Chapter 10.

12.2.4. Discussion of Experiment 4

The detection task of Experiment 2 was repeated using the presentation conditions of Experiment 3 to measure the effect of introducing inter-stimulus gaps on the perception of image identity.

The results indicate that human performance on a stimulus identity detection task is not affected by changing stimulus duration as long as SOA is kept constant. If SOA is reduced, the proportion correct decisions decreases, *vice versa* if SOA is prolonged, the proportion correct decisions increases.

The same is true for the performance of an ideal observer of the activity of single neurones in STSa, although the performance of the ideal observer is lower than that of human observers in the detection task, as expected based on Chapter 10 and the beneficial effect of selective attention in the detection task. Why the human

performance in the detection task is superior to that of single neurones, or what can and what cannot be concluded from comparing single neurones and human behaviour has been discussed in Chapter 10, and the discussion will not be repeated here.

What is important here, is that both (i) the accuracy of human decisions about image identity in RSVP sequence and (ii) the accuracy of the decisions of an ideal observer of the single cell activity in STSa, do not depend on the duration of stimulus presentation as long as SOA is kept constant at 56 or 111ms.

Chapter 10 demonstrated that the performance of an ideal observer of the activity of single neurones in STSa and the performance of human observers react in strikingly similar ways to changes in SOA. This indicated that STSa neurones may participate directly in the performance of the human observers in a task such as the RSVP detection or RSVP memory task. How STSa neurones may may participate in that task remained unclear. In particular, it remained unclear if STSa neurones directly convey some aspect of subjective perception, and if so, what aspect of perception they may contribute.

Experiment 4 demonstrates a surprising fact. RSVP sequences with long gaps seem of poorer quality to a human observers, having less apparent contrast and having much more flicker than their counterpart with 9ms gaps. Based on this difference in perceived stimulus quality, one would assume that both the response of single cells in STSa and the response of human observers in a stimulus identity detection task should suffer from the reduction of perceived stimulus quality obtained by reducing the stimulus duration despite leaving SOA equal. Experiment 4 and the neurometric analysis of the single cell record of Chapter 11 show that these assumptions are mistaken: the performance of both single cells and human observers is *unaffected* by reducing stimulus duration considerably, from 46 to 4ms (Group I) or from 102 to 18ms (Group II), as long as SOA is kept constant. An ideal observer of STSa activity and human subject are unexpectedly linked in their immunity to stimulus duration changes, when it comes to jugging the identity of a stimulus independently of low-level properties.

This finding delivers independent and surprising evidence for the link between STSa neurones and performance in a psychophysical image identity detection task. The fact that the results go against intuition makes the findings more remarkable than the much more predictable finding, that both human perception and the performance of an ideal observer of STSa neurones decreases with decreasing SOA. Together the

results of the present chapter and Chapter 10, supplemented by the findings of Logothetis et al's regarding the relationship between IT and STSa single cell activity and perception under conditions of binocular rivalry (Sheinberg and Logothetis, 1997; Logothetis, 1998), give increasing support for the hypothesis that STSa neurones may directly participate in perception of stimulus identity.

For the reaction times, it was unfortunately impossible to link single cells and human behaviour. Single cells respond earlier in conditions with longer gaps, while human subjects did not. Since human subjects were not encouraged to respond as soon as possible, the reaction times are hard to interpret: it may be that if pressured to respond as fast as possible, human subjects may have shortened their RT in conditions with longer gaps more than in those with shorter gaps. This remains for future experiments to investigate.

What is remarkable about the findings of the present Chapter is that they can help identify the aspects of perception towards which STSa neurones may directly contribute. Experiment 4 measured how well a human subject could detect the *identity* of a stimulus. Single STSa neurones mirrored the performance of human subjects in that task. Hence, in a frame similar to that of Zeki and Bartels (1998), the present findings would be compatible with the idea that STSa may mediate a 'micro-consciousness' of image *identity*: "this is the same stimulus that you showed me before!"[21]. This awareness of identity may be dissociated[3] from the awareness of other aspects of perception, in particular, low-level properties such as flicker.

In the single cell responses, the similarities in conditions with 9 ms gaps and longer gaps allow us to predict, that if other aspects of perception, such as flicker, would also be based on STSa neurone firing, then flicker should be very hard to detect in RSVP sequences even with gaps of 51 and 93ms. Experiment 5 will show that this is not the case, and sharpen our understanding of the potential role of STSa neurones by showing towards which aspects of perception STSa neurones are unlikely to contribute.

---

[21] The single cells recorded in this Chapter are thought to participate in the processing of identity. There is no evidence in the data, that the same cells are sufficient to indicate the former occurrence of the same image. Indeed, the monkey was not performing a matching to sample task during the recording, and hence the single cell recordings cannot be meaningfully analysed to check for neural correlates of working memory for a target stimulus.

12.3. Experiment 5 – Perception of flicker and visible persistence

12.3.1. Introduction to Experiment 5

Experiment 5 was aimed at measuring the smallest gap duration at which subjects start to perceive flicker, and the smallest gap duration at which subjects start to perceive an image as completely vanished from the screen before the next stimulus is presented.

Using the same method as Bowling and Lovegrove (1980) we therefore flashed a single stimulus on and off on the screen for 5 seconds (see Figure 12.2).

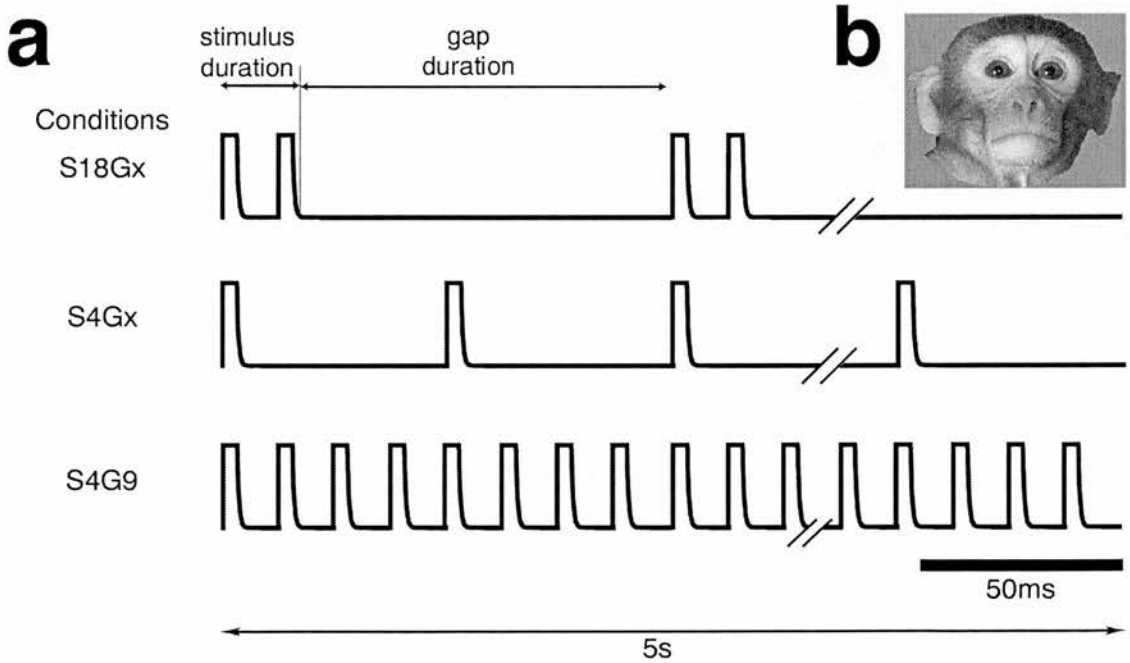


Figure 12.2: Experimental conditions of Experiment 5 illustrated according to the conventions of Fig. 11.2. (a) The single stimulus shown in (b) was presented for either 2 screen frames (i.e. 18ms stimulus duration, top row) or 1 screen frame (i.e. 4ms stimulus duration, middle row) followed by a gap of duration 'x' ranging from 23 to 453ms. In addition, in one condition, the stimulus was presented continuously on the screen, i.e. for 4ms followed by a 9ms gap. The on-off cycle was repeated during 5s. (b) The stimulus used in Experiment 5. The same stimulus had been the best stimulus for cell S154.2, and was the target in one of the 21 stimulus sets of Experiment 2 and 4.

The subjects were asked to indicate verbally after each presentation how they had perceived the sequence. They were instructed to answer "no flicker!" if they had perceived the stimulus as being continuously presented without flicker, "no gap!" if they perceived the stimulus as never completely vanishing from the screen, but flickering, i.e. becoming cyclically brighter and darker or "gap!" if the stimulus was flickering and disappeared completely from the screen between two presentations.

Bowling and Lovegrove (1980) used the same method to flash a grating on and off on a screen, but Bowling and Lovegrove only asked their subjects to decide if the stimulus was perceived to completely disappear from the screen during the gap or not. They defined the visible persistence of a grating as being the longest gap duration between presentations of the grating for which the stimulus was perceived never to fully vanish from the screen. They varied stimulus duration from 50ms to 300ms, and found that shorter stimuli necessitated longer gaps to disappear from the screen. A 50ms stimulus persisted typically for ~250ms, while a 300ms stimulus persisted for ~180ms. This 'inverse duration effect' is typical for visible persistence as indicated in Chapter 3. Bowling and Lovegrove also varied the spatial frequency of the grating, using 1c/deg, 4c/deg and 12c/deg gratings, and found persistence to increase with spatial frequency, with the 12c/deg grating persisted about 50ms longer than the 1c/deg stimulus.

Given that stimulus duration inversely affects stimulus persistence, both the 4ms and 18ms stimulus durations used in Experiment 3 will be used in the present experiment, followed by a variable gap. Also, a typical stimulus from the stimulus set used in experiments 1-4 was used to ensure that the stimulus used to measure persistence and flicker contained a spatial frequency range that was representative for the other experiments of this thesis.

Given that STSa single cell responses are extremely similar for S18G93 and S102G9 conditions and for S4G51 and S46G9 conditions, if STSa activity is the neural correlate of all aspects of visible stimulus perception (a simplistic form of 'grand-mother cell' hypothesis), one would expect subjects to experience the same percept in conditions with similar single cell responses and answer that they perceive neither flicker nor complete stimulus disappearance for gap durations up to 93ms. If on the other hand, as hypothesised above, stimulus *identity* perception is mediated by STSa activity, but low-level properties are mediated by other cortical areas, flicker may be perceived for gap durations much shorter than 93ms. At the same time, if

STSa neurones mediate the perception of stimulus identity, some stimulus presence should be perceived throughout for gap durations up to 93ms, given that STSa neurones continue firing as if the stimulus had stayed on the screen for such gap durations.

According to the idea of micro-consciousness, the percept experienced towards the end of the gap for 93ms gaps may nevertheless be quite different from the percept experienced if the stimulus stayed on the screen, because brain areas mediating the micro-consciousness of low-level stimulus properties (such as flicker etc.), may respond differently from STSa neurones, and show shorter neural persistence. Hence, towards the end of the gap, the perception of identity might be deprived of the normal low-level attributes of the image, rendering perception rather odd.

12.3.2. Methods of Experiment 5

Stimulus and procedure: As indicated in Figure 12.2a, both 4ms stimulus duration and 18ms stimulus duration were tested for the single stimulus shown in Figure 12.2b. Gap durations of 9, 23, 37, 92, 106, 120, 148, 176, 203, 231, 259, 287 and 453ms were tested for both stimulus durations. These gap durations were chosen after preliminary experiments had indicated that subjects stopped reporting "no gap!" and started reporting "gap!" for gap durations of approximately 150-200ms. This resulted in 13 gap durations x 2 stimulus duration = 26 conditions. Each condition was tested 10 times for each subject, in a fully randomised fashion. Each trial began with a 5s presentation of the "stimulus – gap – ... – stimulus – gap" sequence. Subjects were instructed to fixate the sequence, and to report their perception of each sequence whenever they felt ready. Subject were instructed to describe their perception along the following 3 categories: "no flicker", "flicker but no gap", "gap". According to the instructions, subjects answered "no gap" if they perceived the stimulus to be steady on the screen, "flicker but no gap", if they perceived a clear cyclic fluctuation of stimulus contrast over time, but stimulus contrast never reached 0, i.e. the stimulus never disappeared completely from the screen, and "gap", if they perceived the stimulus to completely vanish from the screen before it us turned on again. The experimenter then entered the report into the computer, and started the next sequence. All equipment and room lighting was identical to those used in Experiment 1-4.

A total of 20 trials were obtained for each condition, except for the S4G9 condition, where effectively 40 trials were collected. This was because using a single stimulus, the S4G9 condition is equivalent to the S18G9 condition.

Subjects: David Perrett (DP) and I (CK) participated as subjects in the experiment. Both subjects reported normal vision.

Data analysis: The results of the flicker perception report were so clear-cut, that they necessitated no further analysis.
For the perception of gaps, an ogive function was fitted to the data. For each stimulus duration, the proportion of trials during which no gap was perceived ('y') for any gap duration ('x') was fitted using the following Mathematica® (Wolfram Research, U.S.A.) function: "NonlinearFit[data, 1 - CDF[NormalDistribution[m, s], x], {x}, {m, s}]", where 'data' was the list containing for a particular observer and stimulus duration, the proportion of no gap report for each gap duration, 'm' was the mean, and 's' the standard deviation of the normal distribution whose cumulative probability density function (CDF) was fitted to the data.
The 'gap perception threshold duration' was then defined as the gap duration value for which the fitted ogive crossed the y=0.5 line.

12.3.3. Results of Experiment 5

12.3.3.1. The perception of flicker

For both the 4ms and the 18ms stimulus duration, and for both observers, condition with 9ms gaps were *always* (i.e. 40 out of 40 times) reported as "no flicker", while the conditions with 23ms gaps were *always* reported as "flicker but no gap". Given that using our apparatus it was impossible to vary gap duration in finer steps than entire frame durations (i.e. 14ms), it was not possible to explore at which gap duration flicker was perceived in 50% of the cases.

One of the subjects (DP) differentiated between the 9ms gaps and 23ms gaps, by reporting the 9ms gap sequences as "absolutely no flicker" and the 23ms gap sequences as "no flicker". Considering that he never reported "absolutely no flicker" for 23ms gap conditions, and never reported just "no flicker" for 9ms gaps, it appears

clearly, that he perceived some flicker in the 23 ms condition, but placed his flicker report threshold somewhat lower than I (subject CK; see Figure 12.3 for an tentative illustration of threshold setting). Hence his report "absolutely no flicker" was treated as "no flicker", and his report "no flicker" was treated as meaning "no gap".

Hence for the stimulus used in the present experiment, a 9ms gap is insufficient to perceive flicker when the stimuli are foveated, while a 23ms gap consistently produced a perception of some kind of flicker.

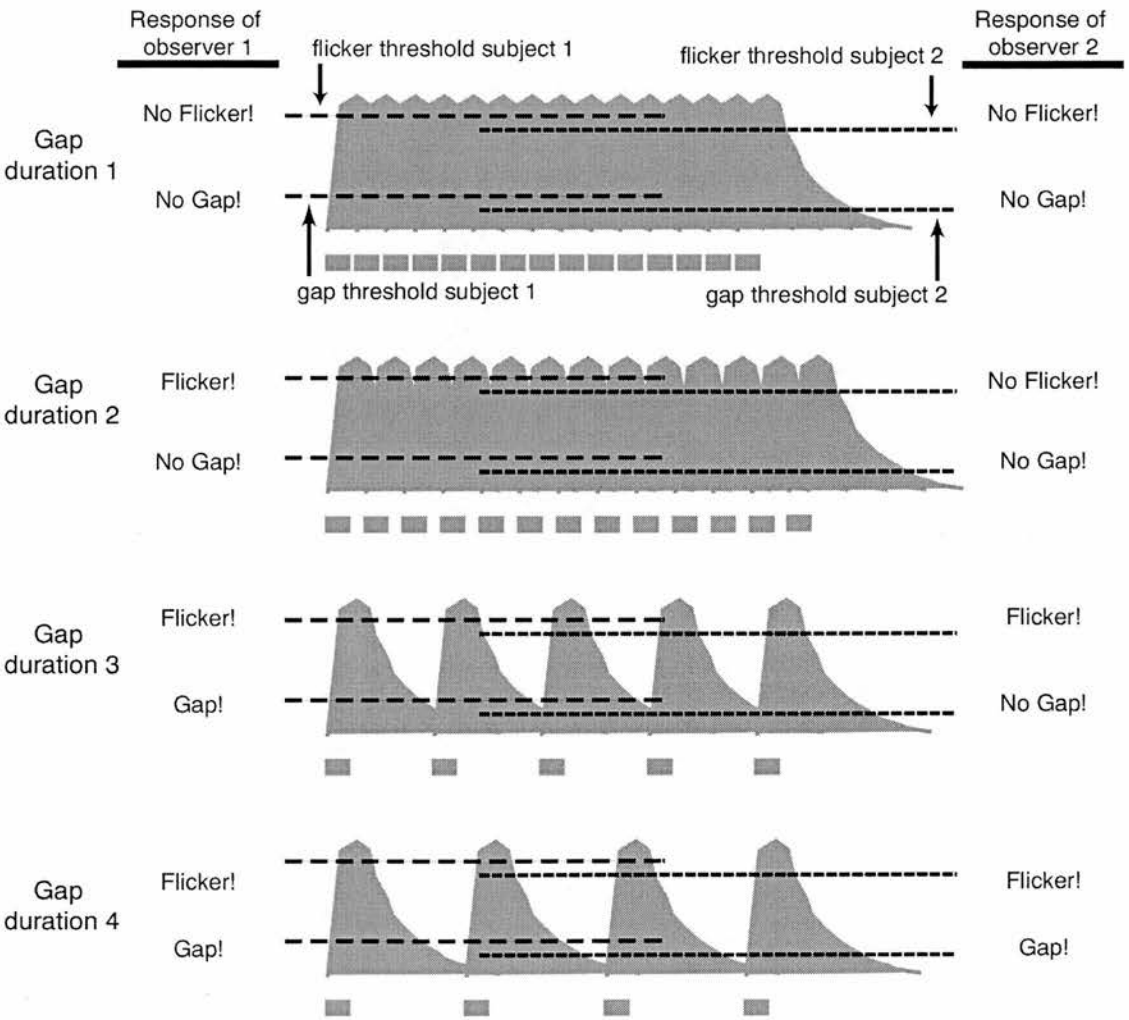


Figure 12.3: Illustration of the concept of individual response thresholds. The y-axis represent a subjective impression of apparent stimulus contrast. The grey areas represent the time course of apparent contrast produced by the brief presentation of a stimulus (as represented by the grey bar under each curve). Four different gap durations are illustrated. The dashed lines represent decision threshold criteria. If the neural response falls under the 'flicker threshold', subjects report flicker, if it falls under the 'gap threshold', subjects report seeing a gap. Threshold for 2 different observers are shown. Observer 1 (left), decided to use higher decision thresholds criteria, and will tend to report flicker and gap perception at shorter gap durations than

observer 2. In the actual experiment observer CK chose a higher thresholds than observer DP.

12.3.3.2. Visible persistence – the perception of gaps

Figure 12.4 illustrates the proportion of trials for which subjects perceived the stimulus as never completely vanishing from the screen. For each subject, and each stimulus duration, an ogive cumulative normal distribution function was fitted to the graph as described in the method section. The point at which this ogive crossed the y=0.5 line was defined as the gap perception threshold duration. Table 12.1 indicates the gap perception threshold duration as a function of observer and stimulus duration. As can be seen, gap perception threshold duration averaged at 175ms for a 4ms stimulus and at 150ms for a 18ms stimulus, displaying the inverse duration effect typical for visible persistence (see Chapter 3).

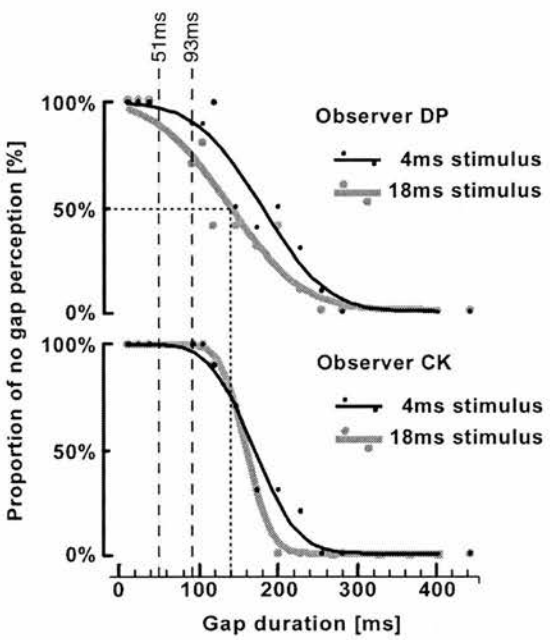


Figure 12.4: The proportion of times in which subjects reported "no gap" is shown as a function of stimulus duration for the two observers separately. "No gap" refers to the stimulus not completely vanishing from the screen. An ogive is fitted to each data set (see Methods). The finely dashed lines represent for observer DP and stimulus duration 4ms, how the gap perception threshold duration is calculated as the gap duration at which the ogive crosses the y=0.5 line. The coarsely dashed lines represent the gap durations used in experiment 3 and 4 and illustrate, how for those gap durations, in a majority of trials, no gaps are perceived.

	S4Gx	S18Gx
DP	180ms	140ms
CK	170ms	160ms
Average	175ms	150ms

Table 12.1: Gap perception threshold duration as a function of observer and stimulus duration. Note the inverse stimulus duration effect.

For a 18ms stimulus followed by a 93ms gap (as used in Experiment 3 and 4), DP perceived gaps in 27% and CK in 0% of the cases. Hence on a majority of trials no gaps are perceived for the S18G93ms condition. The 51ms gap duration used in conjunction with a 4ms stimulus in the S4G51 condition of Experiment 3 and 4, Experiment 5 gives no direct data. Yet, according to the fitted ogive, DP would have perceived gaps in 3% of the cases, and CK in 0% of the cases.

The ogives fitted for CK are steeper than those fitted for DP.

12.3.4. Discussion of Experiment 5

When the image of a face is flashed for 4 or 18ms followed by a gap ranging from 9ms to 453ms, the gap duration at which subjects start to perceive a flicker in the image is very different from the gap duration at which they perceive an image to fully disappear from the screen.

Flicker was never perceived at 9ms gap duration, and always perceived for gap durations of 23ms or longer. Of interest for Experiment 3 and 4 is the fact, that although subjects and single cells responded equally in the S4G51 and S46G9 condition and in the S18G93 and S102G9 condition, Experiment 5 indicates that observers perceive flicker in one but not the other condition. This difference in flicker perception between the conditions contrasts against the lack of difference in the amplitude of STSa responses under those conditions, and supports the idea that STSa is not involved in flicker perception.

Experiment 5 used only a single stimulus in all sequences. To test if flicker is less perceived in the conditions using 8 stimuli, CK rated those conditions with respect to flicker, and clearly perceived flicker in both the S4G51 and S18G93 condition, but not in their S46G9 and S102G9 equal SOA counterparts.

Gap durations much longer than 23ms were necessary for subjects to perceive the image as completely fading from the screen before the occurrence of the next stimulus. Visible persistence, measured as the gap duration for which the ogive crosses the y=0.5 line, i.e. the gap duration at which subjects would report equally often to perceive and not to perceive the stimulus as fully vanished from the screen

before the onset of the next stimulus, was 150ms for 18ms stimulus duration and 175ms for 4ms stimulus duration. At the gap durations used in Experiment 3 and 4, the image was almost never perceived as completely vanishing from the screen. This finding is compatible with the idea that STSa activity, which continues to respond to the stimulus during the entire gap duration, may mediate some visual perception of the stimulus, leading observers to report that the stimulus is still somewhat visible, although it has physically long gone from the screen. The visible persistence of the image may also be due to persisting neuronal activity in other visual areas, in which we did not record.

There were relatively large variations between the results of the two observers: observer CK showed a much sharper transition from no-gap to gap perception than observer DP. This was probably due to observer DP changing his decision threshold for 'completely vanished from the screen' during the experiment. The percept of a stimulus decays after it has been turned off, and it is up to the observer to decide when it completely vanished. In addition, different elements of the stimulus decay at different rates: brighter elements for instance decaying faster (inverse intensity effect, see Chapter 3). Subjects often focused on a particular feature, such as an eye, or an ear, to observe if it fully vanishes. If a subject changes the feature he is paying attention to during the experiment, this will broaden and flatten the ogive curve. Consistent with this interpretation, CK consistently attended to a particular feature of the image and had a steep ogive, while DP reported both changing his criterion and changing the feature he attended to, and had a flatter ogive.

The duration of visible persistence measured in Experiment 5 is in accordance with those found elsewhere in the literature, as described in Chapter 3. In particular the finding of an inverse duration effect supports the validity of the method used in our experiment to measure visible persistence.

Visible persistence brakes down for gaps durations in excess of 300ms. If STSa neurones mediate some aspect of visible persistence, for gap durations in excess of 300ms, veridical responses should become significantly different from persisting responses – a testable hypothesis.

12.4. Conclusions

In Chapter 11, a surprising similarity was demonstrated between the response of STSa neurones to RSVP sequences having the same SOA but very different stimulus durations: introducing relatively long, 51 and 93ms gaps between images, but leaving SOA constant, did not affect the response of STSa neurones.

This surprising finding led to a psychophysical investigation of how different aspects of perception are affected by the introduction of such inter-stimulus gaps.

Experiment 4, using a detection task, demonstrated that the judgement of stimulus identity is also unaffected by stimulus duration as long as SOA is left constant, and this effect is not due to a ceiling effect.

Both human judgement of stimulus identity, and the decision of an ideal observer of STSa single cell activity were shown to be unaffected by gap duration given that SOA is kept constant. Chapter 9 demonstrates that human psychophysical performance and the performance of an ideal observer are also similarly affected by changing SOA. Together, these findings fully support and strengthen the idea that single STSa neurones may be directly involved in the judgement of stimulus *identity*.

Experiment 5 demonstrates that gaps of 23ms create a reliable perception of flicker, while 9ms gaps do not. Gaps of 23ms are substantially shorter than the 51 and 93ms gaps used in the long gap conditions of Experiment 3 and 4. With respect to flicker perception, the conditions S4G51 and S46G9 of Group I and the conditions S18G93 and S102G9 of GroupII are therefore clearly different. Single cells conveying the perception of flicker should thus respond differently in the S4G51 and S18G93 condition compared to the S46G9 and S102G9 condition. STSa neurones do not show such clear flicker perception correlated activity: when comparing the S4G51 and S46G9 condition in particular, the only difference that exists between the population responses to those two conditions is a slight 10ms time shift between responses with the S4G51 condition causing slightly earlier responses. Although not completely impossible, it is very hard to understand how the dramatically different perception of flicker between conditions can be explained by such minimal differences in STAa responses.

Taken together, Experiment 3-5 therefore suggest that neurones in STSa do not mediate all aspects of the visual perception of images in RSVP sequences. It appears that STSa neurones are suitable to directly participate in the judgement of image *identity*, but not in the judgement of *flicker*. Although not explicitly measured in Experiment 5, many other low-level properties of the images are perceived

differently in conditions with long gaps compared to 9ms gaps at the same SOA, including image contrast and brightness. STSa neurones also seem unsuitable to convey those other low-level properties.

As discussed in Section 9.4.4. and Box 9.1. psychophysical investigations only measure behaviour, and not conscious perception per se. Hence, all that can be suggested from Experiment 3-5 is that STSa neurones are suitable for supporting the behavioural response in a psychophysical task requiring subject to judge image identity, but not for supporting the behavioural response in a psychophysical task requiring subjects to judge flicker. Concluding from there, that STSa neurones may be suitable to convey the conscious percept of image identity but not the percept of flicker is left to belief.

Experiment 3-5 are interesting experiments, because they show how both STSa responses, and performance on a stimulus identity judgement are unaffected by stimulus duration. Experiment 3 should be repeated in other visual area. The findings of Levick and Sacks (1970; see my Figure 3.3a) and Duysens et al. (1985; see my Figure 3.2c) in the cat suggest that in the retina and in area 17, responses would not be identical for stimuli of different durations at equal SOA. If the results were the same for monkeys, and if judgement of stimulus identity was based on the response of these neurones, one would predict performance in a psychophysical task requiring the perception of identity to be different for different stimulus duration, even at equal SOA, which was not the case in our experiment. Repeating experiment 3 in different areas along the ventral stream of the macaque monkey would allow identification of which areas show, as STSa, responses that match the psychophysical performance in an identity judgement task, and which areas may match more the performance in a flicker detection task. As discussed in Chapter 10, unless activity in all other brain areas is measured, and STSa is the only area showing a match between single cell responses and perceptual report, one can never know if STSa or an other area is responsible for the performance in the psychophysical task. fMRI investigations may help clarifying this point, by allowing to monitor all areas of the brain simultaneously.

By showing us with what attributes of stimulus perception STSa neurones are associated (i.e. stimulus identity) and of which aspects of perception STSa neurones are dissociated, experiment 3-5 help us specify the perceptual functions of STSa.

Also at present, it is difficult to exclude with certainty that STSa neurones cannot convey some perception of flicker. It is possible, that some aspects of the

neural responses other than response magnitude may be responsible for the perception of flicker – aspects that may have been left unnoticed during our analysis. Also, it may be that the slight differences in response latency, and the higher peak-firing rate in the condition with gaps, particularly in the S18G93 condition may be responsible for the perception of flicker.

Indeed, although the majority of cells in STSa did not show differences in the magnitude of their integrated responses to conditions with long gaps compared to 9ms gaps, 5/21 cells responded differently in the S4G51 condition compared with the S46G9 condition. Yet, those same cells did not differentiate between the S18G93 and the S102G9 condition with respect to response magnitude. It is highly unlikely though that the magnitude difference in the responses of these cells in S4G51 vs. S46G9 was responsible for the perception of flicker, since perception of flicker should then have disappeared in the S18G93 condition, which it did not.

Altogether, although providing no definitive proof, the investigations of the present Chapter suggest that stimulus identity perception is to some degree independent of the perception of other features of the same stimulus such as flicker. In addition, neurones in STSa seem very suitable to participate in identity perception, but rather unsuitable for flicker perception. This favours the idea of a conscious perception as being composed of many micro-consciousness, with STSa participating in a micro-consciousness of identity. As in Chapter 9 and 10, it remains for future experiments to investigate if the link between STSa neurones and psychophysical performance is also a link between neuronal activity and *conscious* perception.

13. Overall Conclusions

The aim of the present thesis has been twofold: first to measure the impact that RSVP presentation rate has on the responses of single cells in the STSa, and second to investigate how the responses of single cells in STSa relates to perception. The first aim derived from the wish to understand better how the brain processes images and to give the research on RSVP a single cell perspective. The second aim derived from a wish to help understand how the brain creates the mind.

Five experiments have been conducted in this thesis: two physiological single cell investigations and three psychophysical investigations. The two physiological investigations (Experiments 1 and 3) measured the responses of single neurones in the STSa while complex images were presented to the monkey in RSVP sequences. The impact of RSVP presentation rate on the responses was measured in Experiment 1, and the impact of replacing part of the stimulus duration with inter-stimulus gaps was measured in Experiment 3. Two of the three psychophysical investigations measured the perceptual counterpart of the single cell responses of the physiological studies: Experiment 2 measured the impact of RSVP presentation rate on human perception and memory of complex images in RSVP sequences while Experiment 4 measured the impact of inter-stimulus gaps on perception. Both of these psychophysical studies investigated the perception of stimulus identity, by asking the question "was this stimulus in the last sequence?" or "tell me when this stimulus appears in the next sequence". The third psychophysical experiment (Experiment 5) measured the impact of interstimulus gaps on a different aspect of perception - the perception of flicker – and did so by asking the question: "did that sequence flicker?" and "did the stimulus disappear?".

In this final chapter, the main conclusions drawn from the results of these 5 experiments will be summarised. All of the conclusions have been discussed previously in the relevant chapters of the thesis. For the benefit of brevity, the references underlying the arguments will not be repeated here, but can be found in the relevant chapters.

13.1. How fast is the visual system?

Experiment 1 demonstrates that the responses of single cells in RSVP can discriminate between complex images presented as rapidly as 72 images/s. Experiment 2 demonstrates that human subjects can detect and memorise stimuli above chance at the same presentation rates. An RSVP rate of 72 images/s corresponds to over thrice the frame rate of cinema motion picture projections.

The majority of the information about the stimulus contained in the spike count of STSa single cell responses is contained in the first 50ms of their response. This is true even for stimuli presented for much longer duration than 50ms: even at 222ms/image, over 50% of the information is contained in the first 50ms of the responses.

The responses of single cells also start to discriminate significantly between stimuli in the first 10-20ms of their responses.

Taken together, these results demonstrate a surprising rapidity of processing in the brain and support the idea that the visual system can process images in a feed-forward processing architecture, with few spikes being transmitted at each stage of processing.

Under normal viewing conditions, where longer processing periods are usually available to process a stimulus, it is likely that feed-forward processing is supplemented by other processing strategies. In particular, the numerous feed-back connections transmitting information from 'higher' visual areas to 'lower' visual areas are likely to play a much more important role if more processing time is available to the visual system. Such supplementary mechanisms could refine the results of visual processing.

Indeed, reducing the SOA in RSVP sequences significantly reduced the information about the stimulus transmitted by the single cell activity. It is possible that the time constraints at the highest RSVP rates excluded processing mechanisms requiring more time, and that depriving the visual system from these supplementary processing mechanisms was responsible for the decrease in transmitted information.

Yet, even in RSVP, feed-back connections could still play a role. Information could travel along faster channels, or shortcut routes to 'higher' visual areas, and feed-

back to 'lower' visual areas in time to have an influence on the first milliseconds of the response of the 'lower' visual areas.

What is surprising is that in many of the RSVP conditions used in this experiment (56-14ms/image), the duration of each stimulus is only a fraction of the response onset latency of the average STSa neurone: e.g. 14ms/image is only ~1/8 of the 108ms average response onset latency measured in the STSa. Under such conditions, more than one mutually exclusive[22] stimulus is likely to be processed simultaneously in the brain. Successful stimulus discrimination at those rates thus requires the architecture of the brain to enable the simultaneous processing of a number of mutually exclusive stimuli. This places limitations in particular on the type of neural code and feed-back connections that can be used in the processing of images.

Yet, the severe decrease in transmitted information as presentation rate is increased suggest that forcing more and more stimuli into the brain simultaneously does result in stimulus competition: the simultaneously represented stimuli reduce each other's response duration and quality. Hence, although the brain appears to be able to process a number of mutually exclusive stimuli simultaneously, it does so at the cost of the quality of processing of individual images. In the present experiments, this trade-off between quantity of stimuli and quality of processing reached an optimum at 42ms/image, where the mutual information per unit of time was maximal (~2bit/s).

13.2. RSVP, neural persistence and stimulus competition

Increasing presentation rate in RSVP sequences shortens the duration of the responses to individual stimuli. For all SOA used in this thesis (222-14ms/image), response duration was equal to SOA plus ~60ms.

Response duration and response amplitude were on the other hand unaffected by replacing large portions of the stimulus duration by inter-stimulus gaps of up to 93ms. This illustrates for the first time in the literature, that in the last stages of the

[22] Mutually exclusive stimuli refers to stimuli that exclude each other in the outside world. For instance a profile and a front view of a face, at the same location on a computer screen are mutually exclusive, in that we never showed 'transparent' faces, and that the screen thus contained one or the other stimulus but never both at the same time.

ventral cortical stream, the neural persistence of a brief stimulus is equal to the response of the same cells if the stimulus had continued to be on the screen – for gaps of up to 93ms.

Taken together, the effect of SOA and the lack of effect of inter-stimulus gaps demonstrate that the concept of stimulus competition applies to conditions of 'temporal stimulus competition' such as RSVP: the representation of a stimulus continues in a ballistic fashion after the termination of the stimulus but only until a new stimulus appears. This new stimulus competes with the former stimulus, reduces the representation of the former stimulus over a period of ~60ms and finally terminates the representation of the former stimulus after ~60ms. The strong neural persistence measured in Experiment 3, and the continued representation of the old stimulus up to 60ms after cells have started responding to the new stimulus (Experiment 1 and 3) are the basis for applying the concept of stimulus competition to cases in which stimuli are not simultaneously presented in the outside world.

On the other hand, increasing the interstimulus gap to 93ms results in a significant increase of the spike count in the first ~60ms of the responses. This effect may be a sign for the competition between the representation of the stimulus of alignment and preceding stimuli. While in STSa, the duration of the representation of the preceding stimulus is independent of the gap duration, in earlier visual areas it is not. In earlier visual cortex, conditions with longer gaps result in less inhibitory interactions between the end of the response to one stimulus and the beginning of the response to the next stimulus. Freed from such interactions, the early components of the responses will be disinhibited by the longer gaps in early visual cortex. This in turn may result in larger early components in STSa.

13.3. Neural correlates of perception

Experiment 1 and 2 show that manipulating the RSVP presentation rate affects the perceptual performance of human subjects and the responses of STSa neurones in similar ways. Moreover, if asked "was this image in the sequence", an ideal observer of the spike count of STSa neurones performs as well, or even better than human

subjects. This demonstrates the correlation[23] between human perception and the responses of single neurones in STSa and that a single neurone can perform as well as an entire human organism when it comes to deciding about the identity of an image. This of course does not mean that the single neurone performing as well as the organism could be the only neurone responsible for the task: a population of neurones will be required. What the findings do show, is that a population of neurones in STSa may thus be sufficient for the perception of the identity of a particular image.

Visual perception in this context is operationally defined, as above chance performance in a psychophysical task requiring the processing of a visual presented stimulus. Whether this perception is conscious or subliminal is a different question that remains unsolved.

Experiment 3 and 4 show that replacing large proportions of a stimulus presentation duration by a blank screen does not affect the perceptual performance of human subjects nor does it affect the performance of an ideal observer of the spike count of STSa neurones. This is true, if single cells and human observers are asked to judge the identity of the stimulus. This lack of effect of gap duration on identity perception is in stark contrast with the strong effect of inter-stimulus gap manipulations on the subjective perception of other visual qualities of the sequences such as flicker (Experiment 5). These findings suggest, that STSa neurones correlate with the perception of stimulus identity but not with the perception of other attributes of the same stimulus such as flicker.

This has strong implications for our concept of how the brain creates the mind: the dissociation of identity perception and flicker perception is incompatible with the idea of a single brain area (STSa) creating the conscious perception of the stimuli. Instead, it supports the idea of visual consciousness as being composed of a number of independent micro-consciousness. Different areas of the brain could contribute to the conscious perception of different properties of an image. STSa could contribute the perception of stimulus identity (e.g. "it is that face!"), while other areas, such as V1 or V2 could possibly contribute to other aspects of perception, such as the perception of flicker. Figure 13.1 illustrates that idea in a cartoon form. Nevertheless a

---

[23] Correlate in this case is not used in the strong, mathematical meaning. Instead, it means that two variables vary in related ways, when an experimental variable is manipulated. In this specific case, it means that both single cell responses and behavioural responses become more accurate when SOA is increased, but as will be seen later, remain identical when SOA is left constant, but gap duration is varied.

central issue – whether consciousness is actually involved in the psychophysical performance – remains unknown, and limits the implications of the present thesis to the behaviourally measurable performance.

While theoretically possible, the idea of a dedicated 'consciousness' areas, which would collect the activity of all other 'unconscious' visual areas (V1, V2, ..., STSa) appears highly uneconomical and unlikely.

What the results from the experiments of this thesis demonstrate is that the responses of a population of STSa cells would be a suitable basis upon which a human organism could base its decision in a psychophysical task regarding stimulus identity. Our introspection tells us, that our responses in such psychophysical tasks are based upon our conscious perception of the stimuli. It is therefore an appealing step to conclude that the activity of STSa neurones *is* our perception of the stimuli. However appealing, it is important to realise that this conclusion, is beyond the realm of empirical science (Box 9.1, Section 9.4.4). Indeed, even the intuition that psychophysical tasks depend on conscious perception has been challenged by findings from blindsight patients and priming experiments in which the performance of subjects shows evidence for the perception of a stimulus while no conscious awareness for the stimulus is reported (see Section 9.4.4). It is up to the reader to decide to believe of not to believe in the rightness of taking that mental step.

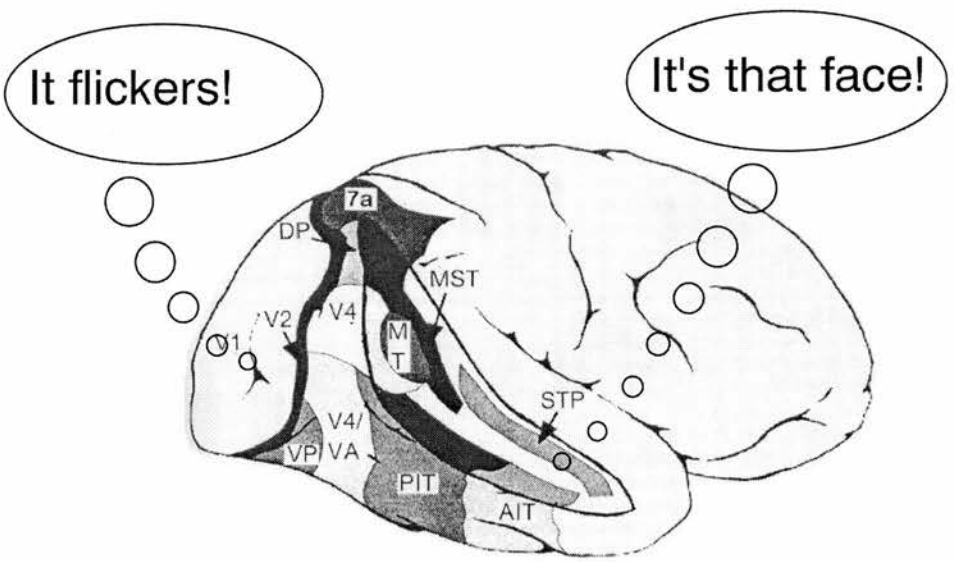


Figure 13.1: The idea of consciousness as composed of many micro-consciousnes. The responses of single neurones in STSa (as measured in Experiment 1 and 3) correlate with the perception of image identity (as measured in Experiment 2 and 4), and would thus be suitable to convey this aspect of perception. Responses in STSa do

not correlate with the perception of flicker (as measured in Experiment 5), suggesting that STSa neurones cannot convey the perception of flicker. Hence, other areas have to be responsible for other aspects of perception, such as flicker.

13.4. Future experiments

From the present thesis, a number of experimental perspective emerge. In particular, it appears that the following experiments could help clarify questions arising from the experiments of the thesis.

First, introducing interstimulus gaps of durations superior to 93ms to identify when the persisting activity to a brief stimulus starts to differ from the activity to a continued stimulus. This experiment is presently being performed.

Second repeating Experiment 3 in other brain areas to identify, which other cortical areas correlate with the perception of stimulus identity as measured in Experiment 4. This would allow us to delimitate cortical areas capable of directly supporting stimulus identity perception.

Third, training a monkey to report its perception of certain stimuli while single cell recordings are made. This would allow measuring the single cell correlates of attentional blink and repetition blindness by quantifying the response of single STSa cells in the period following the detection of a target stimulus. It would also allow a trial-by-trial comparison between single cell responses and the performance of the entire organism.

Forth, experiments exploring the question of how much visual awareness occurred for the correctly reported stimuli in the psychophysical task: was the above chance performance indicating conscious perception or a subliminal process?

Fifth, performing experiments in which new stimuli are used in each trial to explore if the surprisingly high performance of single neurones and human observers in the experiments of this thesis is due to repetition priming or if above chance performance at all rates could also be observed for trial-unique stimuli.

Finally the simultaneous recording of a number of cells with different preferred stimuli while an RSVP sequence containing the preferred stimuli of all those cells would allow to test what the present data can only suggest: are multiple stimuli truly processed simultaneously?

13.5. Concluding remarks

Clearly, we are still at the dawn of the scientific investigation of consciousness. We remain extremely far from an understanding of how single cell activity can create conscious perception or the mind. At present, the very bases of the neural correlates of consciousness remain elusive. In particular, our perception of the world appears unified: we are not awareness of our perception of the world being divided into various micro-consciousnesses. If, as the data of this thesis suggests, the perception of identity may be independent of the perception of other qualities such as flicker, the question arises of how these micro-consciousnesses fuse, or fall into register to create the unified perception we have of the world. Do they naturally fall into register or do they require a binding mechanism to do so? Unfortunately, these intriguing questions remain unanswered.

Also our intuition that to perform correctly in a psychophysical task it is necessary to consciously perceive the stimuli appears to be wrong. The impossibility to directly measure consciousness in an observable way adds to the complication of the field: all we can scientifically investigate are neural correlates of perceptual reports.

What I hope to have achieved with this thesis, is to suggest that perceptual reports may be composed of independent qualities, which may arise from different cell populations located in different brain areas.

By showing that single cells in STSa mirror human performance regarding stimulus identity I hope to have suggested that STSa neurones may play a central role in stimulus identity perception – be it conscious or not.

By showing that an ideal observer of the activity of a single STSa neurone can perform as accurately as a human in a memory task, I hope to have shown that while we still do not know exactly how the brain represents a particular stimulus, this representation includes single cells that are as good at signalling the presence of the stimulus as the entire organism. What this does not mean is that one single cell can be the basis for the psychophysical decision regarding stimulus identity.

At the level of brain physiology, I hope to have provided evidence for how rapid and efficient the primate visual processing can be. I hope these findings will guide our understanding of the mechanisms underlying such rapid visual processing.

Finally by showing how single cells respond during RSVP sequences, I hope to have furthered the understanding of how masking is achieved on a physiological basis: by the competition of otherwise strongly persisting stimuli. Such competition takes about ~60ms for the representation of a new stimulus to replace the representation of an old stimulus. I would like to propose that similar mechanisms may act during binocular rivalry, monocular rivalry, bi-stable perception and masking.

References

Alpern, M. (1953). Metacontrast. *J. Opt. Soc. Am.* **43**: 648-657

Andrews, T. J. and Purves, D. (1997). Similarities in normal and binocularly rivalrous viewing. *Proc Natl Acad Sci U S A* **94**: 9905-8

Andrews, T. J. and Blakemore, C. (1999). Form and motion have independent access to consciousness. *Nat Neurosci* **2**: 405-6

Baker, C. I., Keysers, C. and Perrett, D. I. (1998). Temporal cortex and object permanence: Cells responsive to people occluded from sight. *European Journal of Neuroscience* **10**: 9314

Baker, C. I., Keysers, C., Jellema, T., Wicker, B. and Perrett, D. I. (2000). Coding of spatial position in the superior temporal sulcusof the macaque. *Current Psychology Letters* **1**: 71-87

Bar, M. and Biederman, I. (1998). Subliminal visual priming. *Psychological Science* **9**: 464-469

Baro, J. A., Brzezicki, L. J., Lehmkuhle, S. and Hughes, H. C. (1992). The perceived duration of gratings. *Perception* **21**: 161-6

Bavelier, D. (1994). Repetition blindness between visually different items: the case of pictures and words. *Cognition* **51**: 199-236

Baylis, G. C., Rolls, E. T. and Leonard, C. M. (1985). Selectivity between faces in the responses of a population of neurons in the cortex in the superior temporal sulcus of the monkey. *Brain Res* **342**: 91-102

Bloch, A. M. (1885). Experience sur la vision. *Comptes rendus de seances de la societe de biologie, Paris* **37**: 493-495

Bowen, R. W., Pola, J. and Matin, L. (1974). Visual persistence: effects of flash luminance, duration and energy. *Vision Res* **14**: 295-303

Bowling, A. and Lovegrove, W. (1980). The effect of stimulus duration on the persistence of gratings. *Percept Psychophys* **27**: 574-8

Breitmeyer, B. G. and Ganz, L. (1976). Implications of sustained and transient channels for theories of visual pattern masking, saccadic suppression, and information processing. *Psychol Rev* **83**: 1-36

Breitmeyer, B. G. (1978). Metacontrast masking as a function of mask energy. *Bull. Psychonomic Soc.* **12**: 50-52

Breitmeyer, B. G. and Horman, K. (1981). On the role of stroboscopic motion in metacontrast. *Bull. Psychonomic Soc.* **17**: 29-32

Breitmeyer, B. G. (1984). *Visual masking: an integrative approach.* New York: Oxford University Press

Bridgeman, B. (1975). Correlates of metacontrast in single cells of the cat visual system. *Vision Res* **15**: 91-9

Bridgeman, B. (1980). Temporal response characteristics of cells in monkey striate cortex measured with metacontrast masking and brightness discrimination. *Brain Res* **196**: 347-64

Bridgeman, B. (1998). Durations of stimuli displayed on video display terminals: (n- 1)/f plus persistence. *Psychological Science* **9**: 232-233

Britten, K. H., Shadlen, M. N., Newsome, W. T. and Movshon, J. A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J Neurosci* **12**: 4745-65

Britten, K. H., Newsome, W. T., Shadlen, M. N., Celebrini, S. and Movshon, J. A. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Vis Neurosci* **13**: 87-100

Broadbent, D. E. and Broadbent, M. H. (1987). From detection to identification: response to multiple targets in rapid serial visual presentation. *Percept Psychophys* **42**: 105-13

Brown, M. W. and Xiang, J. Z. (1998). Recognition memory: neuronal substrates of the judgement of prior occurrence. *Prog Neurobiol* **55**: 149-89

Bruce, C., Desimone, R. and Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *J Neurophysiol* **46**: 369-84

Bullier, J. and Nowak, L. G. (1995). Parallel versus serial processing: new vistas on the distributed organization of the visual system. *Curr Opin Neurobiol* **5**: 497-503

Celebrini, S. and Newsome, W. T. (1994). Neuronal and psychophysical sensitivity to motion signals in extrastriate area MST of the macaque monkey. *J Neurosci* **14**: 4109-24

Celebrini, S. and Newsome, W. T. (1995). Microstimulation of extrastriate area MST influences performance on a direction discrimination task. *J Neurophysiol* **73**: 437-48

Chee-Orts, M. N. and Optican, L. M. (1993). Cluster method for analysis of transmitted information in multivariate neuronal data. *Biol Cybern* **69**: 29-35

Chun, M. M. and Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation [see comments]. *J Exp Psychol Hum Percept Perform* **21**: 109-27

Coltheart, M. (1980). Iconic memory and visible persistence. *Percept Psychophys* **27**: 183-228

Coltheart, M. (1980). The persistences of vision. *Philos Trans R Soc Lond B Biol Sci* **290**: 57-69

Coltheart, M. (1983). Iconic memory. *Philos Trans R Soc Lond B Biol Sci* **302**: 283-94

Coltheart, V. (1999). *Fleeting Memories.* Cambridge (MA): MIT Press

Crawford, B. H. (1947). Visual adaptation in relation to brief conditioning stimuli. *Proceedings of the Royal Society of London* **134b**: 283-302

Dehaene, S., Naccache, L., LeClec'H, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., P-F., v. and LeBihan, D. (1998). Imaging unconscious semantic priming. *Nature* **395**: 597-600

Descarte, R. (1824). *Oeuvres de descartes.* VICTOR COUSIN

Desimone, R., Albright, T. D., Gross, C. G. and Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *J Neurosci* **4**: 2051-62

Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philos Trans R Soc Lond B Biol Sci* **353**: 1245-55

di Lollo, V. (1980). Temporal integration in visual memory. *Journal of Experimental Psychology: General* **109**: 75-97

Di Lollo, V. and Bourassa, C. M. (1983). Temporal integration following intensification of long-lasting visual displays. *Vision Res* **23**: 677-87

Di Lollo, V. (1984). On the relationship between stimulus intensity and duration of visible persistence. *J Exp Psychol Hum Percept Perform* **10**: 144-51

Di Lollo, V. and Hogben, J. H. (1985). Suppression of visible persistence. *J Exp Psychol Hum Percept Perform* **11**: 304-16

Di Lollo, V. and Dixon, P. (1988). Two forms of persistence in visual information processing. *J Exp Psychol Hum Percept Perform* **14**: 671-81

Di Lollo, V., Clark, C. D. and Hogben, J. H. (1988). Separating visible persistence from retinal afterimages. *Percept Psychophys* **44**: 363-8

Di Lollo, V. and Dixon, P. (1992). Inverse duration effects in partial report. *J Exp Psychol Hum Percept Perform* **18**: 1089-100

Di Lollo, V., Hogben, J. H. and Dixon, P. (1994). Temporal integration and segregation of brief visual stimuli: patterns of correlation in time. *Percept Psychophys* **55**: 373-86

Di Lollo, V. and Bischof, W. F. (1995). Inverse-intensity effect in duration of visible persistence. *Psychol Bull* **118**: 223-37

Di Lollo, V., Enns, J. T., Yantis, S. and Dechief, L. G. (2000). Response latencies to the onset and offset of visual stimuli. *Percept Psychophys* **62**: 218-25

Diaz-Caneja, E. (1928). Sur l'alternance binoculaire. *Ann. oculist* 721-731

Dick, A. O. (1974). Iconic memory and its relation to perceptual processing and other memory mechanisms. *Percept. Psychophys.* **16**: 575-596

Dixon, N. F. and Hammond, E. J. (1972). The attenuation of visual persistence. *British Journal of Psychology* **63**: 243-254

Dixon, P. and Di Lollo, V. (1991). Effects of display luminance, stimulus meaningfulness, and probe duration on visible and schematic persistence. *Can J Psychol* **45**: 54-74

Duncan, J. (1998). Converging levels of analysis in the cognitive neuroscience of visual attention. *Philos Trans R Soc Lond B Biol Sci* **353**: 1307-17

Duysens, J., Orban, G. A., Cremieux, J. and Maes, H. (1985). Visual cortical correlates of visible persistence. *Vision Res* **25**: 171-8

Engel, G. R. (1970). An investigation of visual responses to brief stereoscopic stimuli. *Quarterly Journal of Experimental Psychology* **22**: 148-166

Enns, J. T., Brehaut, J. C. and Shore, D. I. (1999). The duration of a brief event in the mind's eye. *J Gen Psychol* **126**: 355-72

Eriksen, C. W. and Rohrbaugh, J. (1970). Some factors determining efficiency of selective attention. *Am. J. Physiol.* **83**: 330-342

Fabre-Thorpe, M., Richard, G. and Thorpe, S. J. (1998). Rapid categorization of natural images by rhesus monkeys. *Neuroreport* **9**: 303-8

Fabre-Thorpe, M., Richard, G. and Thorpe, S. J. (1998). On the speed of natural scene categorisation in human and non- human primates. *Cahiers De Psychologie Cognitive-Current Psychology of Cognition* **17**: 791-805

Fehmi, L. G., Adkins, J. W. and Lindsley, D. B. (1969). Electrophysiological correlates of visual perceptual masking in monkeys. *Exp Brain Res* **7**: 299-316

Felleman, D. J. and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex* **1**: 1-47

Felsten, G. and Wasserman, G. S. (1980). Visual masking: mechanisms and theories. *Psychol Bull* **88**: 329-53

Ferrera, V. P., Nealey, T. A. and Maunsell, J. H. (1994). Responses in macaque visual area V4 following inactivation of the parvocellular and magnocellular LGN pathways. *J Neurosci* **14**: 2080-8

Foldiak, P. (1993). The "ideal homunculus": statistical inference from neural population responses. In: *Computation and Neural Systems*, H. Eeckman and J. M. Bower (eds), Norwell (MA): Kluwer Academic Publishers

Fries, P., Roelfsema, P. R., Engel, A. K., Konig, P. and Singer, W. (1997). Synchronization of oscillatory responses in visual cortex correlates with perception in interocular rivalry. *Proc Natl Acad Sci U S A* **94**: 12699-704

Fuster, J. M. and Jervey, J. P. (1981). Inferotemporal neurons distinguish and retain behaviorally relevant features of visual stimuli. *Science* **212**: 952-5

Gawne, T. J., Kjaer, T. W. and Richmond, B. J. (1996). Latency: another potential code for feature binding in striate cortex. *J Neurophysiol* **76**: 1356-60

Georgeson, M. A. and Georgeson, J. M. (1985). On seeing temporal gaps between gratings: a criterion problem for measurement of visible persistence. *Vision Res* **25**: 1729-33

Gershon, E. D., Wiener, M. C., Latham, P. E. and Richmond, B. J. (1998). Coding strategies in monkey V1 and inferior temporal cortices. *J Neurophysiol* **79**: 1135-44

Goodale, M. A. and Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends Neurosci* **15**: 20-5

Green, M. (1981). Spatial frequency effects in masking by light. *Vision Res* **21**: 861-6

Gur, M. and Snodderly, D. M. (1997). A dissociation between brain activity and perception: chromatically opponent cortical neurons signal chromatic flicker that is not perceived. *Vision Res* **37**: 377-82

Haber, R. N. and Standing, L. G. (1970). Direct estimates of the apparent duration of a flash. *Can J Psychol* **24**: 216-229

Heller, J., Hertz, J. A., Kjaer, T. W. and Richmond, B. J. (1995). Information flow and temporal coding in primate pattern vision. *J Comput Neurosci* **2**: 175-93

Hellige, J. B., Walsh, D. A., Lawrence, V. W. and Prasse, M. (1979). Figural relationship effects and mechanisms of visual masking. *J Exp Psychol [Hum Percept]* **5**: 88-100

Hietanen, J. K., Perrett, D. I., Oram, M. W., Benson, P. J. and Dittrich, W. H. (1992). The effects of lighting conditions on responses of cells selective for face views in the macaque temporal cortex. *Exp Brain Res* **89**: 157-71

Hikosaka, K. (1999). Tolerances of responses to visual patterns in neurons of the posterior inferotemporal cortex in the macaque against changing stimulus size and orientation, and deleting patterns. *Behav Brain Res* **100**: 67-76

Holender, D. (1986). Semantic activation without conscious identifi-cation

in dichotic listening, parafoveal vision and visual masking:

a survey and appraisal. *Behavioral Brain Science* **9**: 1-23

Humphreys, G. W. and Bruce, V. (1989). *Visual Cognition.* London: Lawrence Erlbaum Associates Ltd.

Hupe, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P. and Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature* **394**: 784-7

Ito, M., Tamura, H., Fujita, I. and Tanaka, K. (1995). Size and position invariance of neuronal responses in monkey inferotemporal cortex. *Journal of Neurophysiology* **73**: 218-226

Jones, D. S. (1979). *Elementary information theory.* Oxford: Clarendon Press

Judge, S. J., Wurtz, R. H. and Richmond, B. J. (1980). Vision during saccadic eye movements. I. Visual interactions in striate cortex. *J Neurophysiol* **43**: 1133-55

Kahneman, D. (1968). Method, findings, and theory in studies of visual masking. *Psychol Bull* **70**: 404-25

Kanwisher, N. G. (1987). Repetition blindness: type recognition without token individuation. *Cognition* **27**: 117-43

Kanwisher, N. G. and Potter, M. C. (1990). Repetition blindness: levels of processing. *J Exp Psychol Hum Percept Perform* **16**: 30-47

Kanwisher, N. (1991). Repetition blindness and illusory conjunctions: errors in binding visual types with visual tokens. *J Exp Psychol Hum Percept Perform* **17**: 404-21

Kanwisher, N. G. and Yin, C. (1993). Repetition blindness for pictures - evidence for rapid computation of abstract visual descriptions. *Investigative Ophthalmology & Visual Science* **34**: 1131-1131

Keysers, C., Xiao, D.-K., Foldiak, P. and Perrett, D. I. (2000). The speed of sight. *Journal of Cognitive Neuroscience* **(in press)**:

Kinsbourne, M. and Warrington, E. K. (1962a). The effect of an aftercoming random pattern on the perception of brief visual stimuli. *Q. J. Exp Psychol.* **14**: 223-234

Kinsbourne, M. and Warrington, E. K. (1962b). Further studies on the visual masking of brief visual stimuli by a random pattern. *Q. J. Exp Psychol.* **14**: 235-245

Kobatake, E. and Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J Neurophysiol* **71**: 856-67

Kondo, H. and Komatsu, H. (2000). Suppression on neuronal responses by a metacontrast masking stimulus in monkey V4. *Neurosci Res* **36**: 27-33

Kovacs, G., Vogels, R. and Orban, G. A. (1995). Cortical correlate of pattern backward masking. *Proc Natl Acad Sci U S A* **92**: 5587-91

Kovacs, I., Papathomas, T. V., Yang, M. and Feher, A. (1996). When the brain changes its mind: interocular grouping during binocular rivalry. *Proc Natl Acad Sci U S A* **93**: 15508-11

Kratz, K. E. and May, J. G. (1990). Response persistence of cat retinal ganglion cells to the temporally discrete presentation of sinewave gratings. *Int J Neurosci* **52**: 111-9

Lawrence (1971). Two studies of visua search for word targets with controlled rates of presentation. *Perception and psychophysics* **10**: 558-564

Lee, B. B., Martin, P. R. and Valberg, A. (1989). Sensitivity of macaque retinal ganglion-cells to chromatic and luminance flicker. *Journal of Physiology-London* **414**: 223-243

Leopold, D. A. and Logothetis, N. K. (1996). Activity changes in early visual-cortex reflect monkeys percepts during binocular-rivalry. *Nature* **379**: 549-553

Leopold, D. A. and Logothetis, N. K. (1999). Multistable phenomena: changing views in perception. *Trends in Cognitive Sciences* **3**: 254-264

Levick, W. R. and Sacks, J. L. (1970). Responses of the cat retinal ganglion cells to brief flashes of light. *J. Physiol.* **206**: 677-700

Loftus, G. R., Nelson, W. W. and Kallman, H. J. (1983). Differential acquisition rates for different types of information from pictures. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology* **35**: 187-198

Loftus, G. R. and Ginn, M. (1984). Perceptual and conceptual masking of pictures. *Journal of Experimental Psychology-Learning Memory and Cognition* **10**: 435-441

Loftus, G. R., Duncan, J. and Gehrig, P. (1992). On the time course of perceptual information that results from a brief visual presentation. *J Exp Psychol Hum Percept Perform* **18**: 530-49; discussion 550- 61

Loftus, G. R. and Irwin, D. E. (1998). On the relations among different measures of visible and informational persistence. *Cognit Psychol* **35**: 135-99

Loftus, G. R. and Irwin, D. E. (1998). On the relations among different measures of visible and informational persistence. *Cognitive Psychology* **35**: 135-199

Logothetis, N. K. and Schall, J. D. (1989). Neuronal correlates of subjective visual-perception. *Science* **245**: 761-763

Logothetis, N. K. and Pauls, J. (1995). Psychophysical and physiological evidence for viewer-centered object representations in the primate. *Cereb Cortex* **5**: 270-88

Logothetis, N. K., Leopold, D. A. and Sheinberg, D. L. (1996). What is rivaling during binocular-rivalry. *Nature* **380**: 621-624

Logothetis, N. K. (1998). Single units and conscious vision. *Philos Trans R Soc Lond B Biol Sci* **353**: 1801-18

Long, G. M. and Sakitt, B. (1981). Differences between flicker and non-flicker persistence tasks: the effects of luminance and the number of cycles in a grating target. *Vision Res* **21**: 1387-93

Lumer, E. D. and Rees, G. (1999). Covariation of activity in visual and prefrontal cortex associated with subjective visual perception. *Proc Natl Acad Sci U S A* **96**: 1669-73

Macknik, S. L. and Livingstone, M. S. (1998). Neuronal correlates of visibility and invisibility in the primate visual system. *Nature Neuroscience* **1**: 144-149

Macknik, S. L. and Haglund, M. M. (1999). Optical images of visible and invisible percepts in the primary visual cortex of primates. *Proceedings of the National Academy of Sciences of the United States of America* **96**: 15208-15210

Martinez-Conde, S., Macknik, S. L. and Hubel, D. H. (2000). Microsaccadic eye movements and firing of single cells in the striate cortex of macaque monkeys [published erratum appears in Nat Neurosci 2000 Apr;3(4):409]. *Nat Neurosci* **3**: 251-8

Matsumura, M. (1976a). Visual responses to brief flashes of of different temporal stimulus wave forms. *Tohoku Psychologia Folia* **34**: 95-102

Matsumura, M. (1976b). Visual masking by luminance increment and decrement: effect of rise time and decay time. *Tohoku Psychologia Folia* **35**: 104-114

McNicol, D. (1972). *A primer of signal detection theory.* London: George allen & Unwin LTD

Merikle, P. M. (1992). Perception without awareness. Critical issues. *Am Psychol* **47**: 792-5

Milner, A. D., Perrett, D. I., Johnston, R. S., Benson, P. J., Jordan, T. R., Heeley, D. W., Bettucci, D., Mortara, F., Mutani, R., Terazzi, E. and et al. (1991). Perception and action in 'visual form agnosia'. *Brain* **114**: 405-28

Milner, A. D. and Goodale, M. A. (1993). Visual pathways to perception and action. *Prog Brain Res* **95**: 317-37

Milner, A. D. (1997). Vision without knowledge. *Philos Trans R Soc Lond B Biol Sci* **352**: 1249-56

Mishkin, M., Ungerleider, L. G. and Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends in neurosciences* **6**: 414-417

Mitov, D., Vassilev, A. and Manahilov, V. (1981). Transient and sustained masking. *Percept Psychophys* **30**: 205-10

Monahan, J. S. and Steronko, R. J. (1977). Stimulus luminance and dichoptic pattern masking. *Vision Research* **17**: 385-390

Moore, T., Tolias, A. S. and Schiller, P. H. (1998). Visual representations during saccadic eye movements. *Proc Natl Acad Sci U S A* **95**: 8981-4

Morris, J. S., Ohman, A. and Dolan, R. J. (1999). A subcortical pathway to the right amygdala mediating "unseen" fear. *Proc Natl Acad Sci U S A* **96**: 1680-5

Munk, M. H., Nowak, L. G., Girard, P., Chounlamountri, N. and Bullier, J. (1995). Visual latencies in cytochrome oxidase bands of macaque area V2. *Proc Natl Acad Sci U S A* **92**: 988-92

Murata, A., Gallese, V., Kaseda, M. and Sakata, H. (1996). Parietal neurons related to memory-guided hand manipulation. *J Neurophysiol* **75**: 2180-6

Naya, Y., Sakai, K. and Miyashita, Y. (1996). Activity of primate inferotemporal neurons related to a sought target in pair-association task. *Proc Natl Acad Sci U S A* **93**: 2664-9

Neisser, U. (1967). *Cognitive psychology.* New York: Appleton-Century-Crofts

Newsome, W. T., Mikami, A. and Wurtz, R. H. (1986). Motion selectivity in macaque visual cortex. III. Psychophysics and physiology of apparent motion. *J Neurophysiol* **55**: 1340-51

Newsome, W. T., Britten, K. H., Movshon, J. A. and Shadlen, M. N. (1989). Single neurones and the perception of visual motion. In: *Neural mechanisms of visual perception. proceedings of the retina research foundation.*, D. M.-K. Lam and C. D. Gilbert (eds), The Woodlands, Texas: Portfolio Publishing Company

Nickerson, R. S. (1965). Short-term memory for complex meaningful visual configurations: A demonstration of capacity. *Canadian Journal of Psychology* **19**: 155-160

Nowak, L. G., Munk, M. H., Girard, P. and Bullier, J. (1995). Visual latencies in areas V1 and V2 of the macaque monkey. *Vis Neurosci* **12**: 371-84

Optican, L. M. and Richmond, B. J. (1987). Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. III. Information theoretic analysis. *J Neurophysiol* **57**: 162-78

Optican, L. M., Gawne, T. J., Richmond, B. J. and Joseph, P. J. (1991). Unbiased measures of transmitted information and channel capacity from multivariate neuronal data. *Biol Cybern* **65**: 305-10

Orenstein, H. B. and Holding, D. H. (1987). Attentional factors in iconic memory and visible persistence. *Q J Exp Psychol [A]* **39**: 149-66

Panzeri, S. and Treves, A. (1996). Analytical estimates of limited sampling biases in different information measures. *Network-Computation in Neural Systems* **7**: 87-107

Perrett, D. I., Rolls, E. T. and Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. *Experimental Brain Research* **47**: 329-342

Perrett, D. I., Hietanen, J. K., Oram, M. W. and Benson, P. J. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Philosophical Transactions of the Royal Society of London: Series B* **335**: 23-30

Pieron, H. (1925). Recherches experimentale sur la marge de variation du temps de latence de la sensation lumineuse (par une methode de masquage). *L'Annee Psychologique* **26**: 1-30

Potter, M. C. and Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *J Exp Psychol* **81**: 10-5

Potter, M. C. (1976). Short-term conceptual memory for pictures. *J Exp Psychol [Hum Learn]* **2**: 509-22

Raymond, J. E., Shapiro, K. L. and Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: an attentional blink? [see comments]. *J Exp Psychol Hum Percept Perform* **18**: 849-60

Rolls, E. T. (1984). Neurons in the cortex of the temporal lobe and in the amygdala of the monkey with responses selective for faces. *Hum Neurobiol* **3**: 209-22

Rolls, E. T. and Baylis, G. C. (1986). Size and contrast have only small effects on the responses to faces of neurons in the cortex of the superior temporal sulcus of the monkey. *Exp Brain Res* **65**: 38-48

Rolls, E. T. and Tovee, M. J. (1994). Processing speed in the cerebral cortex and the neurophysiology of visual masking. *Proc R Soc Lond B Biol Sci* **257**: 9-15

Rolls, E. T., Tovee, M. J., Purcell, D. G., Stewart, A. L. and Azzopardi, P. (1994). The responses of neurons in the temporal cortex of primates, and face identification and detection. *Exp Brain Res* **101**: 473-84

Rolls, E. T., Tovee, M. J. and Panzeri, S. (1999). The neurophysiology of backward visual masking: information analysis. *J Cogn Neurosci* **11**: 300-11

Sakata, H., Taira, M., Murata, A. and Mine, S. (1995). Neural mechanisms of visual guidance of hand action in the parietal cortex of the monkey. *Cereb Cortex* **5**: 429-38

Salzman, C. D. and Newsome, W. T. (1994). Neural mechanisms for forming a perceptual decision. *Science* **264**: 231-7

Sandell, J. H. and Schiller, P. H. (1982). Effect of cooling area 18 on striate cortex cells in the squirrel monkey. *J Neurophysiol* **48**: 38-48

Sary, G., Vogels, R. and Orban, G. A. (1993). Cue-invariant shape selectivity of macaque inferior temporal neurons. *Science* **260**: 995-7

Schiller, P. H. (1965). Monoptic and Dichoptic visual masking by patterns and flashes. *J exp Psychol* **69**: 193-199

Schiller, P. H. and Smith, M. C. (1965). A comparison of forward and backward masking. *Psychonomic Science* **3**: 77-78

Schiller, P. H. (1966). Forward and backward masking as a function of relative overlap and intensity of test and masking stimuli. *Perception & Psychophysics* **1**: 161-164

Schiller, P. H. (1968). Single unit analysis of backward visual masking and metacontrast in the cat lateral geniculate nucleus. *Vision Res* **8**: 855-66

Schiller, P. H. and Smith, M. C. (1968). Monoptic and dichoptic metacontrast. *Percept. psychophys.* **3**: 237-239

Schmolesky, M. T., Wang, Y., Hanes, D. P., Thompson, K. G., Leutgeb, S., Schall, J. D. and Leventhal, A. G. (1998). Signal timing across the macaque visual system. *J Neurophysiol* **79**: 3272-8

Shadlen, M. N., Britten, K. H., Newsome, W. T. and Movshon, J. A. (1996). A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *J Neurosci* **16**: 1486-510

Shannon, C. E. (1948). A mathematical theory of communication. *AT&T bell laboratories technical journal* **27**: 379-423

Sheinberg, D. L. and Logothetis, N. K. (1997). The role of temporal cortical areas in perceptual organization. *Proceedings of the National Academy of Sciences of the United States of America* **94**: 3408-3413

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning & Verbal Behavior* **6**: 156-163

Shipp, S. and Zeki, S. (1989). The organization of connections between areas V5 and V1 in macaque monkey Visual-Cortex. *European Journal of Neuroscience* **1**: 309-332

Shipp, S. and Zeki, S. (1989). The organization of connections between areas V5 and V2 in macaque monkey Visual-Cortex. *European Journal of Neuroscience* **1**: 333-354

Simons, D. J. (2000). Attentional capture and inattentional blindness. *Trends Cogn Sci* **4**: 147-155

Smythe, L. and Finkel, D. L. (1974). Masking of spatial and identity information from geometric forms by a visual noise field. *Can J Psychol* **28**: 399-408

Snodderly, D. M. and Gur, M. (1995). Organization of striate cortex of alert, trained monkeys (Macaca fascicularis): ongoing activity, stimulus selectivity, and widths of receptive field activating regions. *J Neurophysiol* **74**: 2100-25

Sperling, G. (1960). The information available in brief visual presentation. *Psychological Monographs* **74**: 1-29

Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology* **25**: 207-222

Stoerig, P. and Cowey, A. (1997). Blindsight in man and monkey. *Brain* **120**: 535-59

Stoerig, P., Kleinschmidt, A. and Frahm, J. (1998). No visual responses in denervated V1: high-resolution functional magnetic resonance imaging of a blindsight patient. *Neuroreport* **9**: 21-5

Subramaniam, S., Biederman, I. and Madigan, S. (2000). Accurate identification but no priming and chance recognition memory for pictures in RSVP sequences. *Visual Cognition* **7**: 511-535

Sugase, Y., Yamane, S., Ueno, S. and Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature* **400**: 869-73

Thompson, K. G. and Schall, J. D. (1999). The detection of visual signals by macaque frontal eye field during masking [see comments]. *Nat Neurosci* **2**: 283-8

Thorpe, S. J. (1990). Parallel processing in neural systems. In: (eds), North-Holland: Elsevier

Thorpe, S., Fize, D. and Marlot, C. (1996). Speed of processing in the human visual system. *Nature* **381**: 520-2

Tong, F., Nakayama, K., Vaughan, J. T. and Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron* **21**: 753-9

Tononi, G., Srinivasan, R., Russell, D. P. and Edelman, G. M. (1998). Investigating neural correlates of conscious perception by frequency- tagged neuromagnetic responses. *Proc Natl Acad Sci U S A* **95**: 3198-203

Tovee, M. J., Rolls, E. T., Treves, A. and Bellis, R. P. (1993). Information encoding and the responses of single neurons in the primate temporal visual cortex. *J Neurophysiol* **70**: 640-54

Tovee, M. J. and Rolls, E. T. (1995). Information encoding in short firing rate epochs by single neurones in the primate temporal visual cortex. *Visual cognition* **2**: 35-58

Treue, S. and Martinez Trujillo, J. C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* **399**: 575-9

Treves, A. and Panzeri, S. (1995). The upward bias in measures of information derived from limited data samples. *Neural Computation* **7**: 399-407

Turvey, M. T. (1973). On peripheral and central processes in vision: inferences from an information-processing analysis of masking with patterned stimuli. *Psychol Rev* **80**: 1-52

Ungerleider, L. G. and Mishkin, M. (1982). Two cortical visual systems. In: *Analysis of visual behavior*, D. J. Ingle, M. A. Goodale and R. J. W. Mansfield (eds), Cambridge, MA: MIT Press

Uttal, W. R. (1969a). Masking of alphabetic character recognition by dynamic visual noise (DVN). *Percept. Psychophys.* **6**: 121-127

Uttal, W. R. (1969b). The character in the hole experiment: Interaction of forward and backward masking of alphabetic character recognition by dynamic visual noise (DVN). *Percept. Psychophys.* **6**: 177-181

Volkmann, F. C. (1986). Human visual suppression. *Vision Res* **26**: 1401-16

von Wright, J. M. (1968). Selection in immediate visual memory. *Q. J. Exp. Psychol.* **20**: 62-68

Weiskrantz (1990). *Blindsight.* Oxford: Oxford University Press

Weiskrantz, L. (1996). Blindsight revisited. *Curr Opin Neurobiol* **6**: 215-20

Weisstein, N. (1971). W-Shaped and U-Shaped functions obtained for monoptic and dichoptic disk-disk masking. *Percept Psychophys* **9**: 275-278

Whitten, D. N. and Brown, K. T. (1973). The time courses of late receptor potentials from monkey cones and rods. *Vision Research* **13**: 107-135

Wolfe, J. M. (1984). Reversing ocular dominance and suppression in a single flash. *Vision Res* **24**: 471-8

Wurtz, R. H. (1969a). Response of striate cortex neurones to stimuli during rapid eye movements in the monkey. *Journal of Neurophysiology* **32**: 975-986

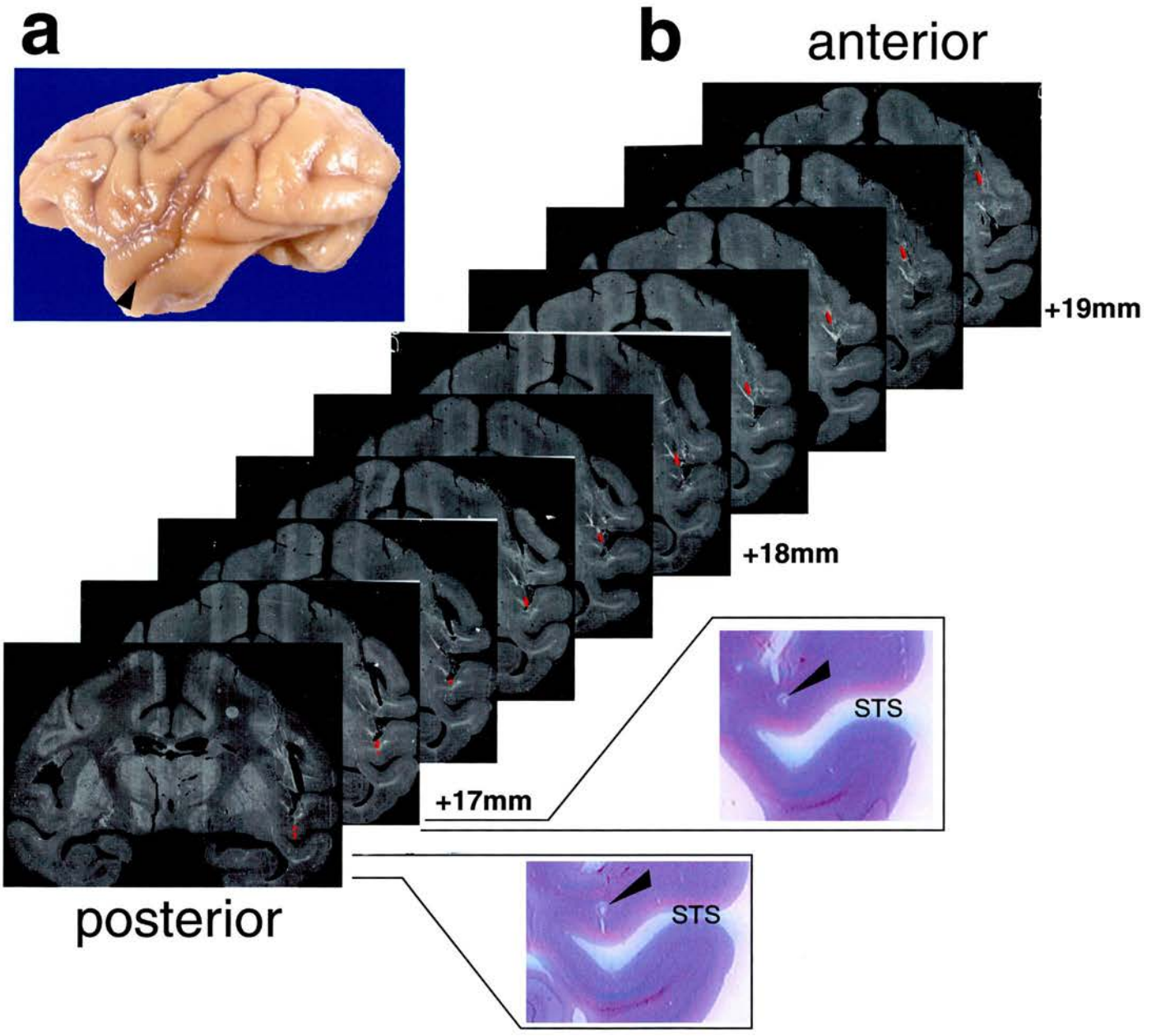
Wurtz, R. H. (1969b). Comparison of effects of eye movements and stimulus movements on striate cortex neurons of the monkey. *Journal of Neurophysiology* **32**: 987-994

Yamane, S., Kaji, S. and Kawano, K. (1988). What facial features activate face neurons in the inferotemporal cortex of the monkey? *Exp Brain Res* **73**: 209-14

Zeki, S. and Bartels, A. (1999). Toward a theory of visual consciousness. *Conscious Cogn* **8**: 225-59

Appendix 1

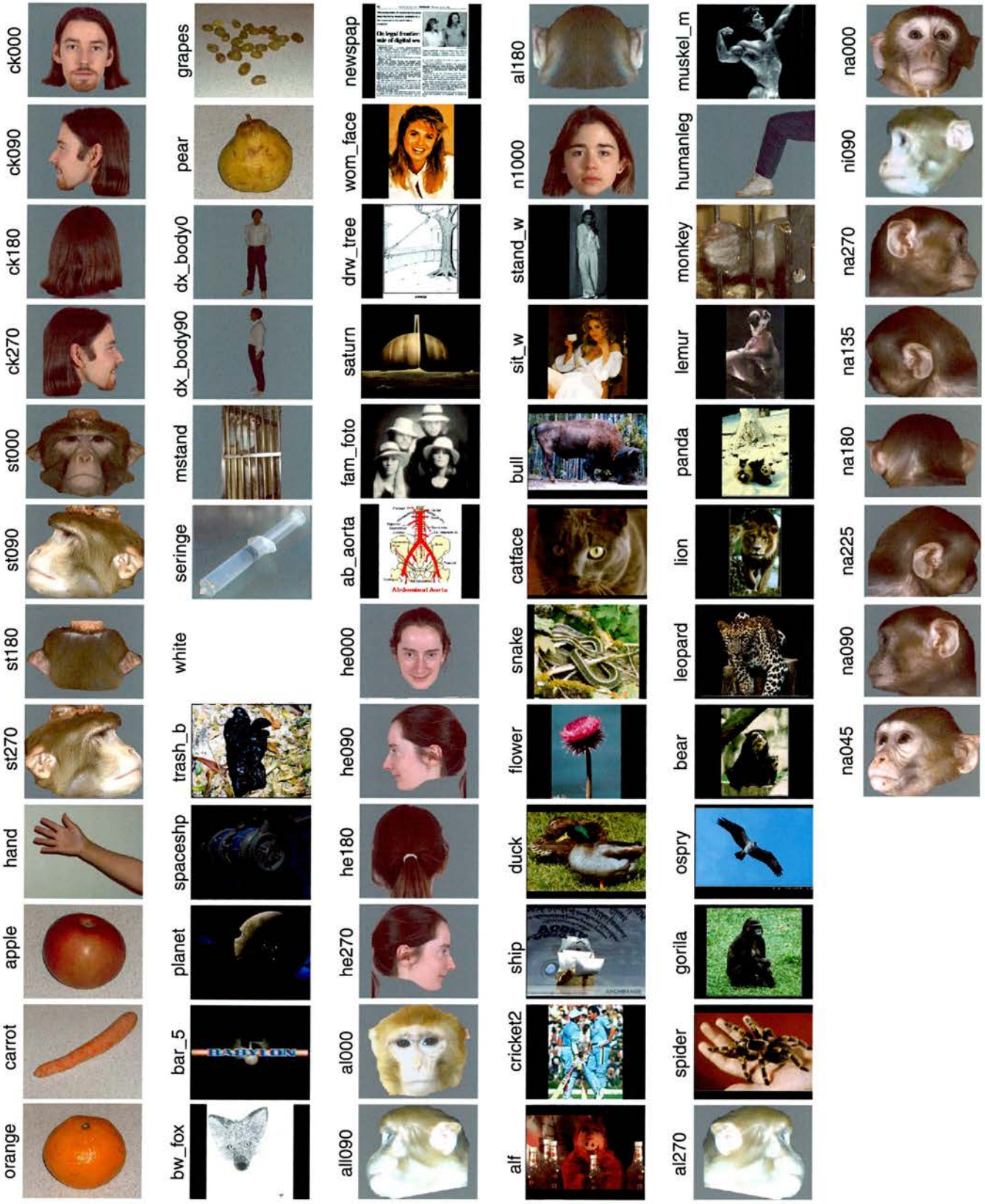
Bellow the Appendix 1 Figure illustrates the histological reconstruction of the final electrode track in monkey "Steve". All other recording positions were reconstructed based on x-rays relative to this reference electrode track.



Appendix 1 Figure: Histological reconstruction of the final recording site. (a) A lateral view of the macaque brain. The black arrow marks the approximate recording position along the length of the superior temporal sulcus (STS). (b) Sequences of 25um coronal sections taken every 250um. The red dots on the right side (i.e. left hemisphere) of each section illustrate the localisations of the fluorescent DiI tracer along the trajectory of the electrode. The electrode track was slightly tilted relative to the coronal plane, explaining why the DiI marking extends over 2mm. The sections have been photographed under white light, and the position of the DiI marking identified using fluorescent microscopy was added onto the photographs. The Anterior-Posterior coordinates of each coronal slice are indicated relative to the interoral plane. On the right, two corresponding, magnified, Nissle stained sections are shown to illustrate the localisation of the electrolytic lesion at the end of the DiI marked electrode track.

Appendix 2

Bellow, the 60 stimuli used in the standard search set plus, in the rightmost column, a set of 8 additional stimuli used for 1 cell showing a particularly clear head direction tuning. For each image, the name of the image is indicated on the left.



Appendix 3: Floating ANOVAs and Bonferroni corrections

When in this thesis a floating ANOVA analysis is performed with one ANOVA p value calculated every millisecond, a total of 550 ANOVAs are typically performed. Given the high number of ANOVAs performed, the significance level, i.e. the Type-I error, becomes difficult to evaluate. To date, there is no exact correction for the Type-I error for multiple comparisons. Bonferroni established that if n statistical tests are performed at individual significance levels of α, the probability of falsely reject one of these n null hypothesis (α'), is less or equal to $n \cdot \alpha$ – the emphasis being on the *less* or equal. Indeed, $\alpha' = n \cdot \alpha$ only if the n tests are performed on statistically independent[24] variables; if the variable are dependent, $\alpha' < n \cdot \alpha$.

In the situation at hand, in which many (550) ANOVAs are performed ms per ms on spike density functions, statistical independence is certainly not granted[25]. First because sdf are the result of a normal convolution of the single cell responses: this creates statistical dependence between adjacent points in time. Second because even before the Gaussian convolution, the state of a neurone at any given time depends on its preceding state. This creates a high degree of dependence between adjacent ANOVA tests and a strict bonferroni correction of the type I error would therefore result in a considerable misestimation of the true Type-I error of the 550 comparisons. The result would be a considerable loss of statistical power and an underestimation of the response duration of the neurone.

In addition, floating ANOVAs are not used in this thesis to claim that for a single ms, the cell responses discriminate between stimuli. Instead, the floating ANOVA demonstrate that many consecutive ANOVAs turn out to have $p<0.05$.

Hence, the true question is what is the probability α'' that chance alone makes k out of n tests statistically significant if all individual tests are performed on

---

[24] Two variables A and B are said to be independent if p(a)=p(a|b).

[25] Statistical independence between the sdf(t) and sdf(t-1) would require, that for any value x and t, p(sdf(t)=x)=p(sdf(t)=x|sdf(t-1)), where sdf(t) is the value of the spike density function at time t, and sdf(t-1) the value of the spike density function at time t-1ms. Clearly, independence is not granted.

statistically dependent variables at a significance level of α=0.05. If the variables had not been dependent, a binomial test could have answered the question. Given the statistical dependence between the variables at hand, the binomial test would also give inaccurate results.

As a results, the approach chosen in this thesis is to use a non-corrected significance level of α=0.05 to determine the duration of the population response. Within this period of population discrimination called the 'window for response analysis', spikes can then be counted to quantify the overall response. These spike counts in turn can be assessed using a single ANOVA to test significant stimulus presentation overall.

This procedure is performed in the awareness that the overall procedure is not performed at α''=0.05, but at an α'' *somewhat* larger than 0.05.