

Visual Attention: Towards a Neural Network Model of Visual Short-Term Memory

Anders Petersen

Kongens Lyngby, 2006
IMM-THESIS-2006-64

Technical University of Denmark
Informatics and Mathematical Modelling
Building 321, DK-2800 Kongens Lyngby, Denmark
Phone +45 45253351, Fax +45 45882673
reception@imm.dtu.dk
www.imm.dtu.dk

Summary

In this thesis a new type of model for Visual Short-Term Memory (VSTM) is presented. The model is inspired by previous work of Usher & Cohen [37], and it links closely with Bundesen's well-established mathematical theory of visual attention, see [5].

We evaluate the model's ability to fit experimental data from a classical type of psychological study, which is known as 'whole and partial report'. Further we compare our results with results obtained using earlier types of models. These models have already successfully assessed the spatial distribution of visual attention; our neural network meets this standard and, on top of that, offers a neural interpretation of how objects are consolidated in VSTM.

In cognitive psychology there are two types of studies have so far been held very isolated from each other. These are studies of Whole Report (WR) and Partial Report (PR) and studies using Rapid Serial Visual Presentation (RSVP).

Bundesen has been very successful in modelling WR & PR data; however no one has so far provided a satisfactory account for phenomena related to RSVP.

It is an interesting finding that our model contains an inherent dynamical behaviour, which potentially could lead to an account for temporally dependent phenomena observed with RSVP.

We hope that in the future, the model will be able to yield a computational account of temporally dependent phenomena like the attentional blink effect, lag-1 sparing, and perhaps even the newly reported cross-over effects for very short inter-stimulus lags ([34]).

Keywords:

Visual Short-Term Memory, Visual Attention, Computational Neuroscience, Cognitive Modelling, Whole and Partial Report, Rapid Serial Visual Presentation, Attentional Dwell-time, lag-1 sparing, Attentional Blink, Theoretical Model

Resumé

I denne rapport præsenteres en ny type model for den visuelle korttidshukommelse. Modellen er inspireret af arbejde udført af Usher og Cohen [37], og den besidder herudover en nær forbindelse til Bundesens veletablerede matematiske teori for visuel opmærksomhed, se [5].

Vi evaluerer modellens evne til at fitte eksperimentelle data fra et klassisk psykologisk studium med rapportering af parallelt præsenterede objekter. Herudover sammenlignes med resultater opnået ved brug af tidligere typer af modeller. Disse har allerede med succes kunne beskrive den rumlige fordeling af visuel opmærksomhed. Vores nye model lever op til de gamle modellens høje standard, og modellen giver samtidig en neural fortolkning af, hvorledes objekter midlertidigt lagres i den visuelle korttidshukommelse.

I den kognitive psykologi har to typer af studier hidtil levet konceptuelt adskilt fra hinanden. De to typer af studier der begge dækker visuel opmærksomhed er: rapportering af parallelt (WR & PR) - samt rapportering af serielt præsenterede objekter (RSVP).

Bundesen har med succes været i stand til at modellere WR & PR data, men til gengæld har ingen hidtil været i stand til at fremkomme med en tilfredsstillende forklaring på fænomener relateret til RSVP.

Det er en interessant opdagelse, at vores model besidder en indbygget dynamisk opførsel, der potentielt kunne lede til et forklaringsgrundlag for tidslige afhængige fænomener opserveret ved RSVP.

Vi håber, at vores model i fremtiden vil føre til en matematisk forklaring på tidsligt afhængige fænomener såsom opmærksomhedsblink, sparring af lag-1 og evt. en forklaring på den nyligt rapporterede 'overkrydsningseffekt' for meget korte interstimulus intervaller ([34]).

Preface

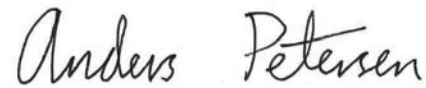
This Master's thesis was carried out at the Technical University of Denmark, Department of Informatics and Mathematical Modelling in the period from September 2005 to July 2006.

The thesis accounts for a work-load of 40 ECTS points.

Supervisor on the project was Professor Lars Kai Hansen. Primary co-supervisor was Associate Professor Søren Kyllingsbæk and secondary co-supervisor was Assistant Professor Henrik Aanæs.

Parts of the thesis are to appear in the paper: "Visual Attention – A Neural Network Model of the Visual Short-Term Memory, which has been submitted to the Neural Information Processing Systems Conference, 2006 (NIPS2006). The paper can be found in Appendix C.

Kgs. Lyngby, July 2006



Anders Petersen

Anders Petersen, s001737

Acknowledgements

First of all I wish to thank my supervisors for great inspiration and keen interest in the work related to this master thesis.

Further, I want to thank Claus Bundesen for access to the experimental data in [38], and also I shall thank Marius Usher for kindly verifying parameters and settings in [37].

I also thank Mads Peter Sørensen for a very inspiring discussion related to the Fokker-Planck method for obtaining solutions to stochastic differential equations.

I would also like to thank my fellow students at IMM; in part for good company and in part for helping me out with various IT related questions.

Finally I warmly thank everyone who spent time reviewing this report. In my opinion, there is no doubt that you helped me improve the quality of my writing by a factor of two!

Thank you all..!

Contents

Summary.....	i
Resumé	iii
Preface	v
Acknowledgements	vii
Contents.....	viii
1 Introduction.....	1
1.1 Visual Attention.....	1
1.2 Visual Short-Term Memory	1
1.3 Modelling VSTM	2
1.4 Brain versus Computer	3
1.5 Statement of hypotheses	4
1.6 Report overview	4
PART I: THEORY – REVIEWING VISUAL SHORT-TERM MEMORY	7
2 Visual Short-Term Memory – an overview	9
2.1 Theory of Visual Attention.....	9
2.2 Input and output – processes related to TVA	11
2.3 The input side – Object matching.....	11
2.4 The output side – the Attentional Race to enter VSTM	12
2.5 Limited processing capacity	13
2.6 Limited storage capacity.....	13
2.7 A theory for masking.....	14
2.8 A theory for re-distribution of processing resources.....	15
2.9 Spike VSTM models	16
3 Psychological experiments probing the Visual Short-Term Memory	19
3.1 Modelling psychological experiments.....	19
3.2 Whole and Partial Report Experiments	19
3.3 Attentional dwell time Experiments	21
PART II: METHODS – REVIEWING VSTM MODELS	25
4 Fitting the storage capacity of VSTM in a whole report study.....	27
4.1 Fitting with the binomial distribution.....	27
4.2 Fitting with the hypergeometric distribution.....	28

5	Preliminary testing of Usher & Cohen’s original model	29
5.1	The model	29
5.2	Implementation	30
5.3	Critique of Usher & Cohen’s model	32
6	The two-component mixture model (2CMM and 2CMM ₂)	35
6.1	Shibuya & Bundesen’s original version	35
6.2	Optimization method	36
6.3	Critique of Shibuya & Bundesen’s optimization method	37
7	The multi-component mixture model (MCMM)	39
7.1	An improved mixture model	39
7.2	Optimization method	40
8	The unit spike model (USM)	41
8.1	Modifying Usher & Cohen’s model	41
8.2	Adopting the parameters: C , t_0 and α from Shibuya & Bundesen	42
8.3	Validity of simulation and model parameters	44
8.4	Shibuya & Bundesen’s data – Estimating parameters for the model	44
8.5	Dynamics of the model	46
9	The non-unit spike model (NUSM)	49
9.1	Modifying the unit spike VSTM model	49
9.2	Shibuya & Bundesen’s data – Estimating parameters for the model	49
9.3	Dynamics of the model	50
10	The conservatory non-unit spike model (CNUSM)	53
10.1	Modifying the non-unit spike VSTM model	53
10.2	Shibuya & Bundesen’s data – Estimating parameters for the model	54
10.3	Attentional dwell time modelling	56
10.4	Dynamics of the model	59
PART III: RESULTS – FIT TO DATA		63
11	Results	65
11.1	Fitting the storage capacity of VSTM in a whole report study	65
11.2	Preliminary testing of Usher & Cohen’s original model	66
11.3	The fit to Shibuya & Bundesen’s partial and whole report data	67
11.4	The fit to attentional dwell time data	69
PART IV: DISCUSSION, CONCLUSION AND FUTURE WORK		71
12	Discussion	73

12.1	Opening up the black box of visual attention	73
12.2	Experimental design: Fokker-Planck versus Monte-Carlo.....	73
12.3	Binomial and hypergeometric fits to whole report data	74
12.4	Mixture model fits to whole and partial report data	75
12.5	Spike model fits to WR & PR and to attentional dwell time data.....	75
13	Conclusion	79
14	Future work.....	81
Appendix A	FIRM equations – whole and partial report.....	82
	Parameters in FIRM.....	82
	Probability distributions.....	83
	Probabilities	84
Appendix B	Simulink diagrams.....	90
	The GrandScheme, ver. 1.....	90
	The GrandScheme, ver. 2.....	90
	The NTVA Block.....	91
	The Spike VSTM block	91
	The repmat block	92
	The Euler Integrator block.....	93
	The F block	93
	The G block.....	93
	The H block.....	93
Appendix C	Paper for NIPS2006.....	94
Nomenclature.....		102
	Abbreviations.....	102
	Symbols.....	103
References		104

1 Introduction

1.1 Visual Attention

Although the brain is very potent when it comes to processing of information, the system can only achieve an optimal result, if irrelevant information can be discarded so as not to take up vital computational resources.

In our everyday life it is of utmost importance for us that we are able to perceive, comprehend, and react to what ever impressions we receive. Often, like when driving a car, our rate of success is heavily dependent upon how efficient and how fast we can process, interpret and react to sensory stimuli.

Considering the visual domain much processing takes place already in the early stages before the information has actually reached the brain. In the following we shall refer to ‘visual attention’ as the process, influenced by higher (cortical) areas in the brain, which enables us to focus processing resources to certain ‘selected objects’ among the variety of all objects in the visual scene. We assume as does the Theory of Visual Attention [5] that objects that have been selected are characterized by being the ones ‘encoded’, i.e. the ones entering visual short-term memory (VSTM).

1.2 Visual Short-Term Memory

What does it mean that a visual categorization of a certain object enters VSTM? We refer to the neural theory of visual attention ([3] p. 302), which follows the ideas of Hebb (1949) among others. We assume that when a visual categorization of an object enters the VSTM, this can be understood as a population of neurons (representing the categorization) having their activation sustained by being incorporated into a type of feedback loop.

We now define the concept of Visual Short-Term Memory (VSTM).

We consider VSTM as a mechanism temporarily capable of holding on to a limited number of selected visual objects. The object that we consider is either a digit or a letter entity.

In relation to the work by Miller (1956) what we consider as objects can also be defined as ‘chunks’ of visual information [29], hence in principle a word or a number consisting of more than one digit could also count as a visual object.

If not representation of the visual objects is upheld by either continued visual presentation or executive control from higher cortical areas, memories will simply

decay away inside a temporal span of less than a second. This is consistent with Sperling's (1960) report of the very fragile so-called iconic memory [39].

Following Philips (1971) iconic memories are fragile and decay rapidly, whereas visual short-term memories can be robust to subsequent stimuli and last for a period of many seconds [33].

Cattell already in the late 19th century demonstrated a surprising limit in how many objects that can be perceived at the same time – a limit only about 4 objects which may be held in the VSTM at the same time [7],[10]. This finding is independent of the number of objects visually presented [39].

Evidence further exist that the “magical number” of 3-to-4 objects is largely independent of how many features that are encoded for each object. This means that the complexity of the visual object does not hold an influence on the memorial capacity of the VSTM (see [28], but see also [1]). This latter empirical finding is consistent with Bundesen's Theory of Visual Attention [5], which assumes that VSTM is limited only by the number of visual objects rather than by the number of features that these objects contain.

1.3 Modelling VSTM

Modelling the function of the VSTM it is essential that the inherent capacity limitation is properly mimicked. Most likely the VSTM would be heavily overloaded, if the system should lack the ability to represent only the most salient of the visually appearing objects.

For the above mentioned reason a proper understanding of the mechanism behind visual attention seems vital in establishing an appropriate model for VSTM. Visual attention has been explored in experimental studies such as WR & PR and RSVP, and it seems quite clear that a good model should possess the ability to fit data from these kinds of experimental studies.

The more objects there are in the visual scene, the more unlikely it is that attention is drawn to a specific predefined object. This is because in principle all objects have a chance of becoming encoded into VSTM. According to Bundesen all objects in the visual scene take a place in what one could think of as a ‘race’ to become encoded. In Bundesen's race model [38], p. 597 the ‘odds’ that a given object is selected as a winner in the race is directly related to the rate value with which the object participates.

When it comes to modelling WR & PR data from multi-element visual displays, we have no knowledge of any better VSTM model than Bundesen's race model, at least not at present.

1.4 Brain versus Computer

The human brain contains in the order of 100 billion neurons, which is a very large number compared to the number of transistors in a modern computer processor. (In 2006 a newer version of the Pentium 4 processor contains approximately 178 million transistors).

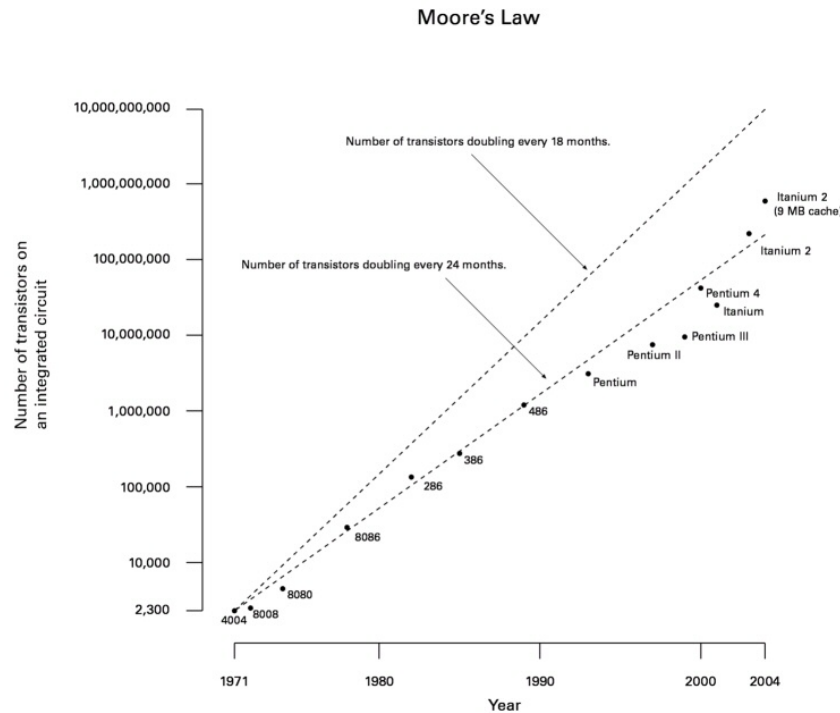


Figure 1: Growth of transistor counts for Intel processors (dots) and Moore's Law (upper line: doubling in 18 months; lower line: doubling in 24 months). [32]. Having proved its validity in the last 35 years Moore's law predicts that the number of transistors in a standard processor will however exceed the number of neurons in the human brain; and this only in a matter of less than 20 years! Even foreseeing that computer processors should one day contain as many transistors as the human brain contains neurons, it is most unlikely that this fact alone would generate the basis for artificial intelligence.

In fact there are vast differences between the structure of the brain and the structure of a modern-day computer. The computational steps in a computer take place very fast compared with the brain, but the brain on the other hand takes advantage of a massive parallelism due to the fact that neurons are highly inter-connected compared to the units of the processors – the transistors.

The high inter-connectivity is very much related to the fact that the brain is a 3-dimensional structure, whereas a normal chip is merely 2-dimensional. Typically each neuron is connected to thousands of other neurons allowing for a completely different computational paradigm than the one, which is used by modern day computers. In this thesis we wish to explore visual attention, which is perhaps one of the most important cognitive processes implemented in the brain. Visual attention is closely linked with the visual short-term memory, which can be considered as sort of the working memory (that is the Random Access Memory, RAM) of the visual part of the brain.

1.5 Statement of hypotheses

This assumption behind this thesis is that visual short-term memory can be fitted with appropriately chosen mathematical functions. We model data from three different types of experimental studies: 1) whole report, 2) partial report and 3) attentional dwell-time studies.

1) In the case of whole report studies, we hypothesize that the distribution of the score j (correctly reported targets, averaged over a number of trials) can be modelled with a simple statistical distribution. We suggest that the binomial or even the hypergeometric distribution would be suitable distributions in this respect.

2) In the case of partial report, we hypothesize that the model used by Shibuya and Bundesen (1988) can in a rather simple way be improved to cover also the outliers of the data set; yielding a better fit for the complete set of whole and partial report data in [38].

Also in relation to partial report, we hypothesize that a winners-take-all kind of neural network (inspired by Usher & Cohen's model in [37]) can in a feasible fashion be integrated with Bundesen's (1990) theory of visual attention. This neural network type of model, we hypothesize, will be able to yield a closer fit to the data [38]. Especially we hope the model will be superior to Shibuya & Bundesen's model, when it comes to modelling the distribution of the scores j

3) Based on the fact that the neural network type model we propose works in a dynamical way, we speculate that this will provide an option to model temporally dependent aspects of visual attention, especially we hope to be able to learn more about the neurobiological mechanism behind the famous 'attentional blink effect'.

1.6 Report overview

The report is organized in four parts, which we shall now present to the reader.

PART I: THEORY – REVIEWING VISUAL SHORT-TERM MEMORY

In chapter 2 the theory of visual attention is presented. Here we inspect relevant cognitive processes related to visual attention. We explain some of the limitations of the short-term memory and we explain also the concept of masking etc. We further present a couple of figures that illustrates how we developed Bundesen's theory of visual attention to include a more sophisticated model for visual short-term memory.

In chapter 3 we review some of the most important visual attention experiments. These psychological experiments serve for us to obtain a better understanding of the behaviour of the visual short-term memory. In latter chapters of the report we shall model exactly such experiments.

**PART II:
METHODS – REVIEWING VSTM MODELS**

In chapter 4 we use a binomial and hypergeometric distribution to fit a whole report study.

In chapter 5 we conduct a preliminary testing of Usher & Cohen's STM model, which is to a great extent the model on which we base our new VSTM model.

In chapters 6 to 10 we present a number of models that we use to Shibuya and Bundesen's comprehensive set of WR & PR data.

In chapter 6 we review Bundesen's original mixture model based on the FIRM equations.

In chapter 7 we implement an improved version of Bundesen's mixture model.

In chapter 8 we present the first version of our new neural network VSTM model, namely the unit-spike model.

In chapter 9 we present the non-unit spike model, which is developed from the model in chapter 8.

In chapter 10 we present the conservatory non-unit spike model, which is in a similar way developed from the model in chapter 9.

**PART III:
RESULTS – FIT TO DATA**

In chapter 11 we present results from fitting the whole report study, as described in chapter 4. Further we present the results from the preliminary testing of Usher & Cohen's model in chapter 5. Further we present a table listing the results from fitting the complete whole and partial report study of Shibuya and Bundesen with the models mentioned in chapters 6 to 10.

**PART IV:
DISCUSSION, CONCLUSION
AND FUTURE WORK**

In chapter 12 we discuss the various results that we found. We list the main contributions in chapter 13, and further we propose a few topics for future work in chapter 14.

Finally, we also included three appendices, which contain respectively: the 'FIRM equations – whole and partial report', the Simulink diagrams and the Paper for NIPS2006.

Nomenclature and References can be found in the last pages of the report.

PART I:
THEORY – REVIEWING
VISUAL SHORT-TERM MEMORY

2 Visual Short-Term Memory – an overview

2.1 Theory of Visual Attention

The Theory of Visual Attention (TVA) developed by Bundesen [5] is a unified theory of visual recognition and attentional selection. The TVA model assumes that objects in the visual scene have previously been segregated. Each object is characterized by its combination of more or less present or even absent features. These features could be as primitive as different colors or textures, but they could also relate to quite complex features such as advanced descriptions of shape. An important assumption of the TVA-model is that ‘feature extraction’ has already taken place. Therefore we assume that the degree of alignment with the various features is well-known for all objects in the visual field; we say that “visual evidence” is already at hand.

TVA provides a mathematical framework describing how the visual system is able to select individual objects in the visual field, based on the visual evidence and the setting of two different types of visual preference parameters. These parameters represent the influence from higher cortical areas including the Visual Long Term Memory (VLTm).

The output of the TVA model is a set of rate parameters v that are directly related to the probability that a given characterization, *object x belongs to category i* , is encoded into the short-term memory. Following [5] the rate parameters are given by:

$$v(x, i) = \eta(x, i) \beta_i \frac{w_x}{\sum_{z \in S} w_z} \quad (2.1)$$

Where

$$w_x = \sum_{j \in R} \eta(x, j) \pi_j \quad (2.2)$$

Here $\eta(x, i)$ is defined as the strength of the sensory evidence that object x belongs to the visual category i . The pertinence of the visual category¹ j is denoted by π_j and setting of these values effectively implements the so-called *filtering* mechanism. The perceptual decision bias of a visual category i denoted by β_i . Setting of these values conversely implements a complementary mechanism called *pigeonholing*.

¹ Please note that we generally apply j for the score in WR & PR, i.e. the number of reported targets. Here however j is used to denote a visual category, this is consistent with Bundesen’s choice of symbols when he formulated the theory of visual attention in [5].

The filtering mechanism increases the likelihood that elements belonging to a target category are perceived; this without biasing perception in favor of perceiving the elements as belonging to any particular category.

Pigeonholing conversely changes the probability that a particular category i is selected without affecting the conditional probability that element x is selected given that category i is selected.

For example if subjects are to report the identity of red target digits amongst green distractor digits, then the green distractors must be filtered out, and the targets must be categorized with respect to digit identity. In TVA terms pertinence should be high for red and low for green stimuli and bias should be high for digit identities and low for other categories e.g. letters.

A neural interpretation of TVA is given in [3]. Basically here pigeonholing (selection of features) is considered as an increase in the rate of firing of neurons, while filtering (selection of objects) is considered as an increased mobilization of neurons.

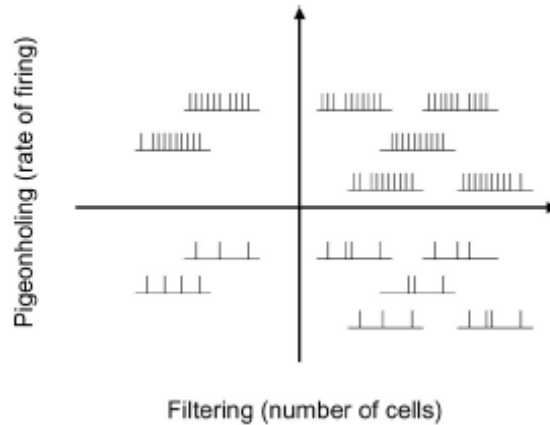


Figure 2: Spike train illustration of the two complementary mechanisms in visual attentional tasks, namely pigeonholing and filtering. Adopted from [3] with permission.

Corresponding to the interpretation in NTVA the fraction: $w_x / \sum_{z \in S} w_z$ in equation (2.1), which is the relative attentional weight of object x compared to the weight of all objects z in the visual field S , can be directly interpreted as the relative fraction of neurons mobilized to process a given object x , compared to the total number of neurons processing any given object z belonging to the visual field S .

We define the effective exposure duration: $\tau = t_e - t_0$

Here t_0 should be considered a temporal threshold before conscious processing begins, and this is a standard assumption in many experimental settings. A typical value for healthy subjects is $t_0 \approx 20$ ms.

The density p as a function of time τ for an exponential random process with rate parameter ν is defined as:

$$p(\tau) = \nu e^{-\nu\tau} \quad (2.3)$$

The corresponding distribution function P found by integration is:

$$P(\tau) = \int_0^{\tau} p(t) dt = 1 - e^{-\nu\tau} \quad (2.4)$$

Provided we assume that the visual short-term memory has infinite capacity, the function $p(\tau)$ can be defined as the probability density that a characterization with rate parameter ν is stored in VSTM exactly at time τ . The distribution function is simply the accumulation of probability density up until time τ , which indicates that the $P(\tau)$ can be defined as the probability that the characterization with rate parameter ν has been stored in VSTM at time τ .

2.2 Input and output – processes related to TVA

In the periphery of Bundesen's theory of visual attention we find two important cognitive processes. On the input side TVA assumes that *Object Matching (I)* has been carried out. Closer to the output side is the *Attentional Race (II)*, where the 'strength' of each participant is directly determined from TVA.

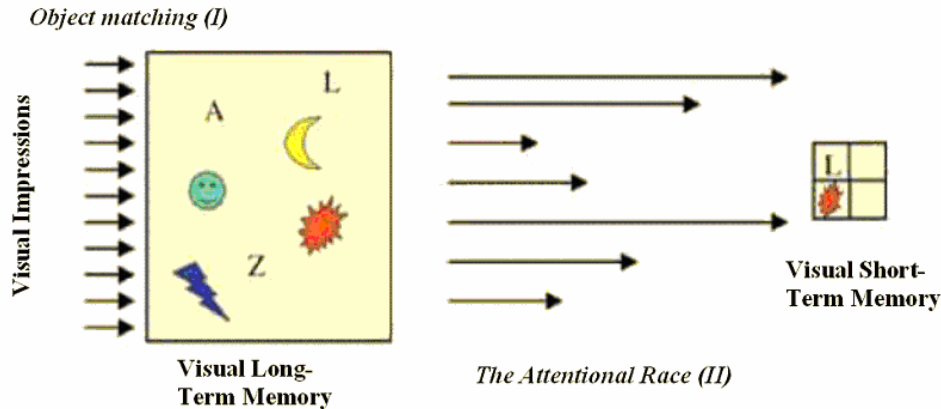


Figure 3: Bundesen's TVA interpretation of visual attention. Note the two important cognitive processes i.e. *Object Matching* and *The Attentional Race*. Note also that here the VSTM is modelled simply as a container with a fixed capacity of four objects.

2.3 The input side – Object matching

The first process is simply gathering of visual evidence. We speak of this process as '*object matching*' (see Figure 3), since we find it relevant to think that objects in the visual field are to some extent 'matched' against object representations in the VLTM. In this thesis we do not consider the problem concerning which feature extraction techniques are biologically plausible or even technically most optimal to use.

In TVA the visual evidence $\eta(x, i)$ is defined as the strength of the sensory evidence that object x belongs to the visual category i . In fact TVA does not consider how these values are calculated, though it seems evident that objects have already been segregated from the visual context in which they appear.

Adding our own interpretation to this matter, we can assume that the visual field has been split up into a background region (which we do not consider) and further the object regions. The latter belong to the visually presented objects.

We now assume that the object regions are ‘matched’ against some feature templates. In a very rough interpretation we can simply imagine that each alpha-numeric character has its own template, and that all the visual objects are somehow matched against this set of templates.

In reality calculation of visual evidence of course goes on in a much more sophisticated way, which it is beyond the scope of this thesis to consider.

Here we only note that receptive fields in the primary visual cortex can be modeled quite accurately by 2D Gabor functions (See e.g. [13], p. 62). As an example Deco and Zihl (2004) used features, which were extracted by use of Gabor filters, to model the phenomenon of visual neglect [15]. The computational model they applied, quite interestingly, shares many common features with the STM model of Usher & Cohen [37].

Finally we note that lately also Independent Component Analysis (ICA) has been used to model the response of simple V1 receptive fields; Hyvärinen and others have even taken the first steps to use ICA to model V2 complex receptive fields [23].

Three bullet points serve as important milestones in obtaining a more complete understanding of how visual evidence is computationally established:

- Understanding the process of segregation
- Understanding how receptive fields extract visual features
- Understanding the hierarchy in which features are arranged

Again we emphasize our assumption that visual evidence is somehow available a priori.

2.4 The output side – the Attentional Race to enter VSTM

The second process that we shall review is ‘*the attentional race*’. (See Figure 3 above) According to Shibuya & Bundesen (1988) every object in the visual scene takes a place in what one could think of as a race to become encoded into VSTM. In Shibuya & Bundesen’s race model the ‘odds’ that a given object is selected as a winner in the race is directly related to the rate value v with which the object participates. It is worth noting that the race is a stochastic rather than a deterministic process meaning that no one can beforehand predict readily, which objects will win the race.

The attentional race effectively decides which objects get stored in VSTM, which is a process also known as consolidation of visual short-term memories. Selecting which objects that should enter the VSTM is the overall aim of visual attention; hence understanding how the VSTM works of course plays a very important role, if the theory of visual attention should be improved.

2.5 Limited processing capacity

Following Bundesen (1990) we assume that the processing capacity C of the VSTM is limited. Further we interpret C as the total processing rate summed over the set R of all visual categories i and further summed over all objects x in the visual field S .

$$C = \sum_{x \in S} \sum_{i \in R} v(x, i) \quad (2.5)$$

Assuming that the display contains only homogenous elements (cf. Bundesen, 1990) in the sense that there exists a constant k , such that for every element x in any one of the displays:

$$\sum_{i \in R} \eta(x, i) \beta_i = k \quad (2.6)$$

Then by substitution of equation (2.1) and equation (2.6) into equation (2.5) we see that:

$$C = \sum_{x \in S} \sum_{i \in R} v(x, i) = \sum_{x \in S} \sum_{i \in R} \eta(x, i) \beta_i \frac{w_x}{\sum_{z \in S} w_z} = \sum_{x \in S} k \frac{w_x}{\sum_{z \in S} w_z} = k \frac{\sum_{x \in S} w_x}{\sum_{z \in S} w_z} = k \quad (2.7)$$

Thus across the set of displays considered, the processing capacity C is fixed at k .

Throughout the experiments mentioned in this paper we shall always assume that we are dealing with homogenous elements, i.e. we assume that the processing capacity C is constant. Homogeneity can be approximated by ascertaining that visual elements are of equal size, and that lighting and stimulus/background contrast is constant all over the visual display.

2.6 Limited storage capacity

According to [38], besides being limited by processing capacity C , the short-term-memory also must deal with a limited-storage capacity K . Hence in [38], p. 524 it is suggested that an encoding will only take place, if the storage capacity is not already fully occupied due to previously encoded categorizations.

In the literature it has been widely discussed whether the number of ‘spaces’ available in memory is actually 3, 4 or perhaps 5 spaces. Whether this ‘magic number’, as it has been referred, should actually be understood in such a strict and rigid way is probably

doubtful. In [5] the ‘magic number’ is set to K . Bundesen also underlines that the ‘magic number’ is subject-dependent.

The model in [5] is constructed so (given that $K = 3.74$) that it assumes a mixture model, i.e. 74 % of the times the memory equals 4, and the rest (i.e. 26 %) of the times the memory conversely only equals 3. This interpretation seems simple, but it is also very rigid, especially because a typical finding is that visual memory capacity seems more broadly distributed, see Figure 4.

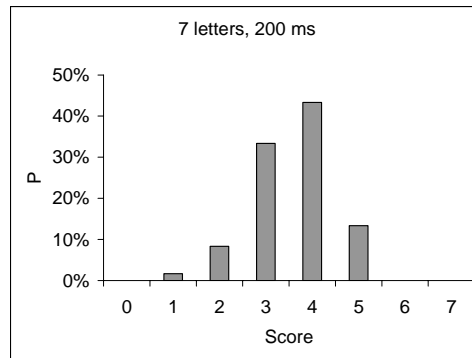


Figure 4: A typical score distribution. The histogram was obtained for a subject with normal visual acuity. A total of 60 trials were conducted. Data were kindly provided by Kyllingsbæk.

As we see the score distribution in Figure 4 has some similarity with the binomial distribution. As an alternative distribution Kyllingsbæk suggests that a closer fit to the memorial capacity empirically found can be obtained with a hypergeometric distribution.

We believe that a suitable model should in some way be able to mimic the true score distribution, and our intention is to discover whether a plausible VSTM model can be determined, which fulfills this criterion.

2.7 A theory for masking

Masking is a technique applied in a wide span of different experiments assessing visual attention. In a sense the principle is fairly simple. Without masking, if a subject is exposed to a visual stimulus, the chance that the subject will successfully be able to report the identity of the stimulus is virtually not dependent on how long time the stimulus was presented. With masking however, if a subject is exposed to a visual stimulus, the frequency distribution (the proportion of times the subject is able to correctly report the identity of the stimulus), is heavily dependent upon the exposure duration, i.e. how long time the stimulus was presented.

In this thesis our hypothesis always is that the effect of a mask is simply ‘abrupt termination of stimulus processing’, i.e. from the moment the mask has been presented, the stimulus is offered no new chances of being encoded into VSTM. (This is equal to saying that the probability density of entering VSTM drops immediately to zero at the onset of the mask.)

Again we shall only consider WR & PR, when the stimulus display is terminated by masks. Further we only consider the case where the rate of processing $v(x,i)$ is constant in time within the effective exposure duration. In future work with unmasked stimulus display, we could for instance assume that the v values decay exponentially with a fixed time constant μ . In formal terms and without masking we would have:

$$v(x,i;t) = \begin{cases} 0 & \text{for } t \leq t_0 \\ \eta(x,i)\beta_i \frac{w_x}{\sum_{z \in S} w_z} & \text{for } t_0 \leq t \leq t_e \\ \eta(x,i)\beta_i \frac{w_x}{\sum_{z \in S} w_z} \exp(-(t-t_e)/\mu) & \text{for } t > t_e \end{cases} \quad (2.8)$$

Here t_e is the exposure duration. Parameter μ varies between individuals, and it also depends on the experimental design (e.g., contrast and intensity of the display). It is typically estimated in the range of a few hundred milliseconds. Given the dynamical nature of the spike models it is an interesting topic for future work to investigate implications of non-masked elements in the visual display, especially because in the rather common RSVP experiments masking is not generally used.

2.8 A theory for re-distribution of processing resources

Consider serial presentation of two visual elements as illustrated in Figure 5 below:

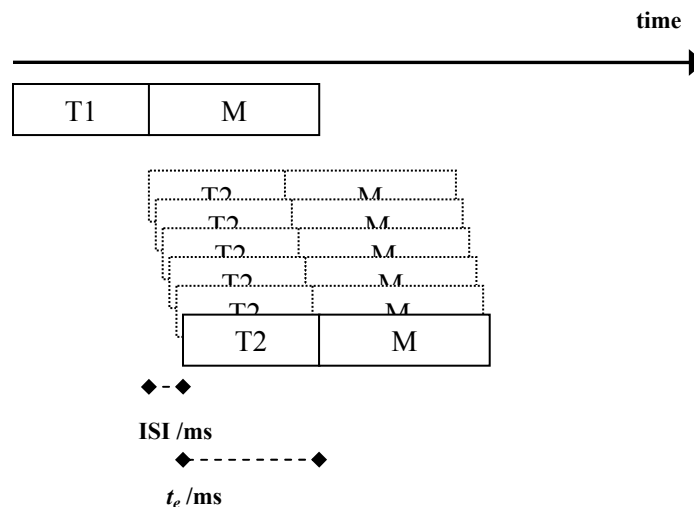


Figure 5: Serial presentations of two masked targets with very short inter-stimulus intervals (ISIs).

According to equation (2.5) all processing capacity C is initially assigned to the processing of target T1. We assume that the mask M does not attract any processing resources,² hence in principal it should be suspected that all processing capacity is also available to the processing of target T2, simply due to the fact that the two targets do not overlap in time. We always assume that the processing capacity is able to re-distribute itself instantaneously. This means that we assume that the encoding rate for target T1, v_{T1} and the encoding rate for target T2, v_{T2} are constant and equal to C at all times.

In future work it could be considered, if a more realistic description, of how processing capacity is re-distributed, could be found. In section 2.7 we assumed that an exponential decay term could be used to model the effect of masking; possibly a very similar approach could be taken in order to model temporal dynamics related to re-distribution of processing resources. Again we emphasize that in this thesis we always assume processing capacity to be instantaneously re-distributed and further determined according to equation (2.5).

2.9 Spike VSTM models

A promising STM model was presented by Usher & Cohen (1999)³. Inspired by this model we propose a new type of model, namely the spike VSTM model. A motivation for this type of model is that it links directly with several important assumptions expressed in Bundesen's Theory of Visual Attention [5]:

- TVA assumes that encoding rates v , that are calculated from the visual evidence (η -values) are available; the basis for this is the setting of the mental set (β - and π -values). The rates v can be used directly as inputs in our new model.
- TVA assumes that objects take a place in race to become encoded into VSTM. Such a race can actually be directly implemented in the proposed model.
- TVA assumes a finite storage capacity K , however the capacity is also assumed to vary from trial to trial. Again the latter assumption fits perfectly with the behaviour of our new model. This we know from Usher & Cohen (1999).

In Figure 3 above we noted that the VSTM was modelled simply as a container with a fixed capacity K of 4 objects. The spike VSTM model does not have a fixed capacity; here the score j is the outcome of a winners-take-all type of competition between the visual objects. With the spike VSTM model we no longer consider VSTM a static store. This leads us to Figure 6, which is more or less a revision of Figure 3.

² This is probably not completely correct. Perhaps, rather than neglecting the mask, it is more correct to count the mask as sort of a distractor. This issue however is left for future work.

³ An alternative relevant model was proposed by Raffone & Wolters in [35]. This model appears however to be considerably slower to update. Compared to Usher & Cohen's model Raffone & Wolters' model does not implement a mean field approach to the activation level of the neuron assemblies, rather the dynamics in the model is based on the implementation of a classical (often slow) spiking neural network.

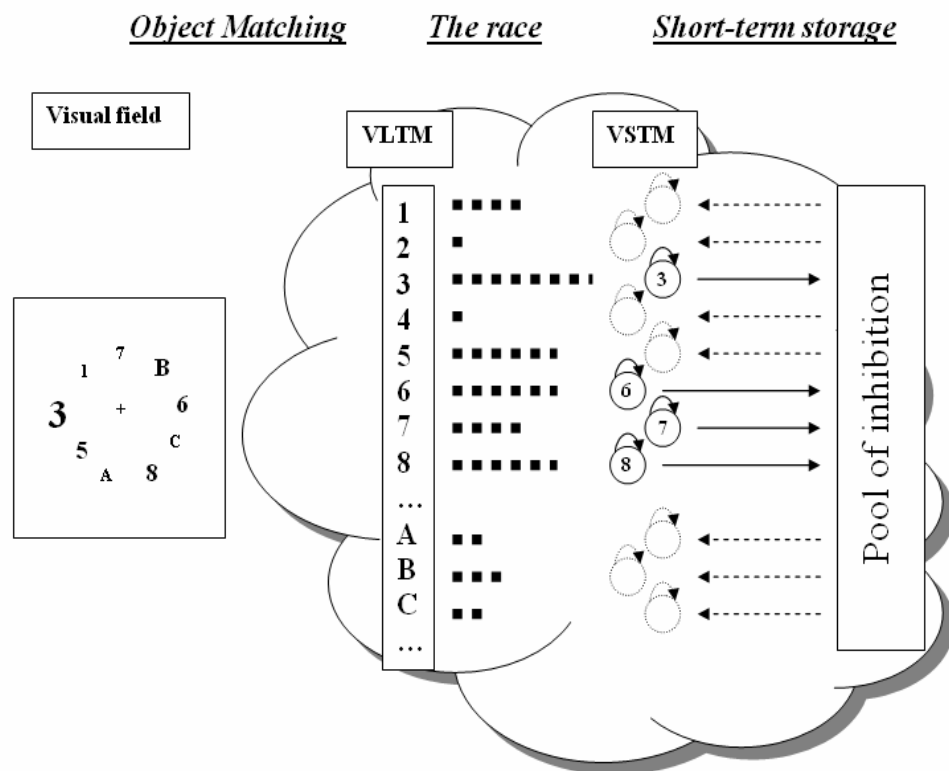


Figure 6: Revised interpretation of visual attention. Now VSTM is modelled as a winners-take-all kind of network. The example depicts an inhomogeneous (see section 2.5) partial report experiment. Among the targets (i.e. digits) those that occupy a spatially larger area are more likely to enter VSTM. In principle however also smaller ones can enter VSTM, as can also the distractors (i.e. the letters). The figure depicts a typical situation, where four objects have succeeded in consolidating themselves in VSTM.

3 Psychological experiments probing the Visual Short-Term Memory

3.1 *Modelling psychological experiments*

In this chapter two different kinds of psychological experiments will be reviewed. First in section 3.2 we explain the concept of whole and partial report studies, and later in section 3.3 we explain the concept of attentional dwell time studies. It is these two kinds of experimental studies that we shall seek to model in later chapters of the report. The purpose of this chapter is to give the reader a chance to become familiar with the concepts behind the two kinds of experimental psychological studies. Note that both studies aim at exploring visual attention, and hence both studies potentially provide important information when it comes to modelling the visual short-term memory.

3.2 *Whole and Partial Report Experiments*

In the classical whole report experiment the subject is asked to report all displayed elements. All elements are presented simultaneously and for an equal amount of time.

In partial report the experimental setup is very similar. The only difference is that now the elements are divided into a group of targets and a group of distractors, and the user is asked only to report the targets.

Typically and as is the case in the experiments that we shall review, all displayed elements are followed by a mask. In our definition a mask is a visual element that covers the part of the visual field, where the original element was initially displayed.

The purpose of the mask is to stop the originally presented element from continuing their 'fight' to become encoded into VSTM. If the mask was not present, a 'sensory after image' would be able to survive, despite the fact that the originally displayed element was in fact removed from the display. (Cf. [39].)

During the combined whole and partial report experiment in [38] the subject was presented with a visual field consisting of a number of digits and letters. The subject was asked to report digits only.

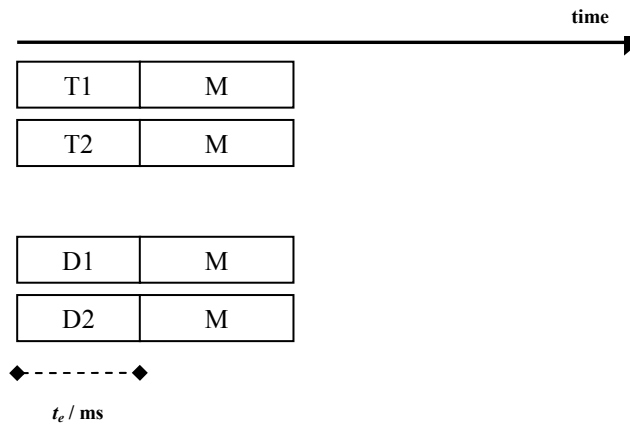


Figure 7: A partial report experiment in which two targets, T1 and T2 and two distractors, D1 and D2 are presented simultaneously. These are all followed by a visual mask, M that effectively serves to cut off further processing of the originally presented element.

The data of Shibuya & Bundesen (1988) covers the performance of a subject, MP participating in an extensive series of whole and partial report experiments. The subject was instructed to report targets, i.e. digits while ignoring distractors, i.e. letters displayed on an imaginary circle around a small fixation cross at the centre of the screen. Experimental trials covered twelve different combinations of total number of 2 – 6 targets, T , and 0 – 6 distractors, D . Further, exposure durations t_e were varied systematically at 10, 20, 30, 40, 50, 70, 100, 150 and 200 ms. Each experimental condition was repeated 60 times but trials were mixed so that the subject had no a-priori knowledge of the experimental condition. Moreover trials were grouped into blocks to minimize the element of fatigue. Each presented character was immediately followed by a mask lasting for 500 ms. Further information can be found in [38].

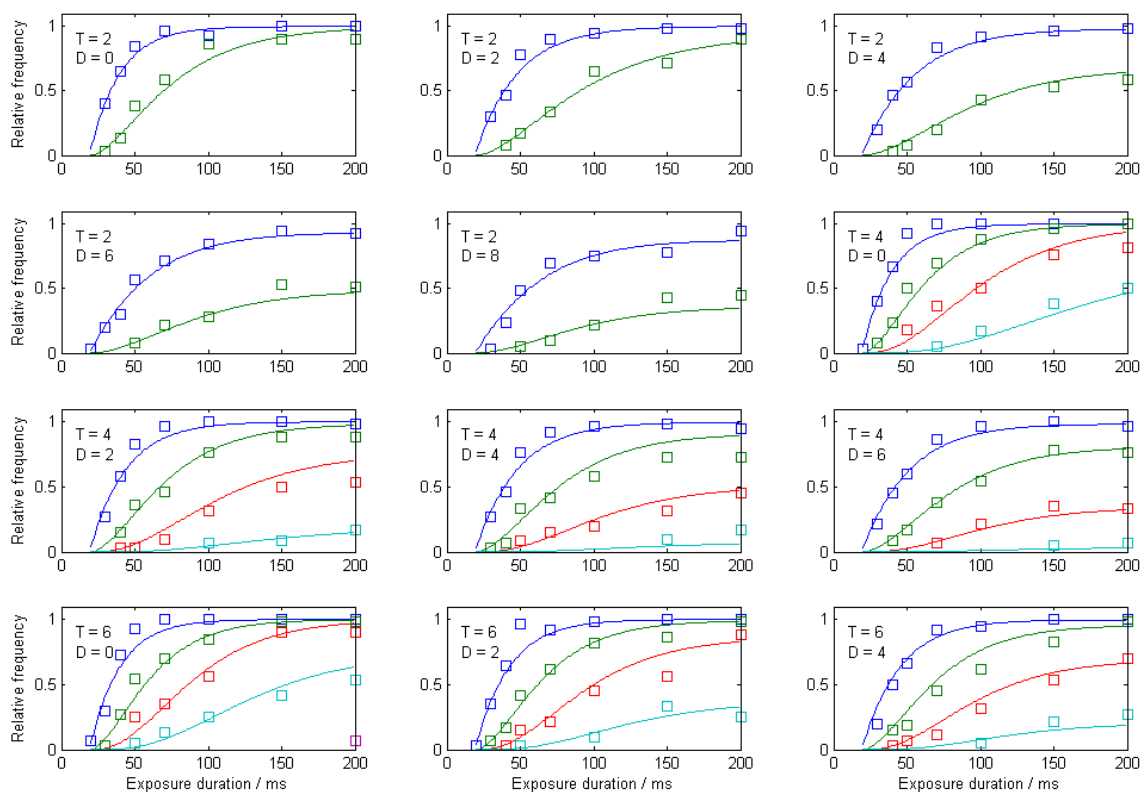


Figure 8: Relative frequency of scores of j or more (correctly reported targets) showed as a function of exposure duration. The inter-plot varied parameters are the number of targets T and the number of distractors D , while the intra-plot varied parameter is j . The solid curves represent the theoretical fit provided by Shibuya & Bundesen’s original 2-component mixture model. (See section 6.1) For clarity observed frequencies less than 0.02 are omitted from the figure. The figure is a reconstruction of Figure 4 in [38] – with the same settings as given in [38], p. 598. (Subject MP)

3.3 Attentional dwell time Experiments

In the classical RSVP experiment a stream consisting of single visual elements (chunks) is presented serially in time. The stream traditionally contains many elements, both targets and distractors, which are presented at a fixed time interval. Usually an interval of 100 ms is used between presentations. The first input presented is termed the lag-0 input, the input presented after 100 ms is termed the lag-1 input, the input presented after 200 ms is termed the lag-2 input, and so forth.

Considering visual attention as a finite resource it is interesting to note that the allocation of attention to a temporally leading object often implicates a reduction in report performance on trailing objects. It is well-known that the identification rate of subjects that are asked to report two serially displayed characters is in fact heavily dependent on the Stimulus Onset Asynchrony (SOA). This phenomenon is usually referred to as the attentional blink. A typical result from an attentional blink study is shown in Figure 9:

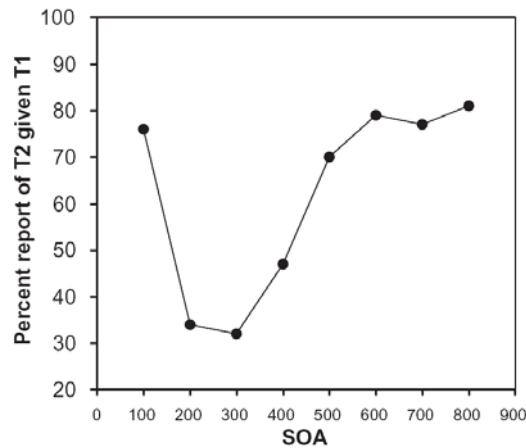


Figure 9: A characteristic attentional blink result in which performance on target 2 (T2) is reported as a function of SOA, conditional on correct report of target 1 (T1). Targets were letters, and distractors were digits. Figure adopted from [8].

Ward, Duncan & Shapiro (1996) used a slightly modified version of RSVP, where only two visual elements were presented in series. (See Figure 10) Their objective was to measure the duration of the attentional dwell time, i.e. for how long the firstly presented element would interfere with the second element presented.

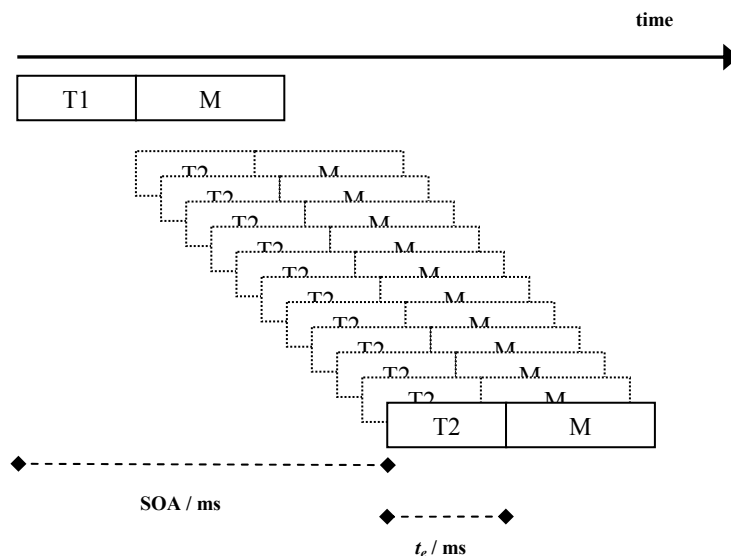


Figure 10: Experimental setup in [40]. The SOA between the two visual elements T1 and T2 is the parameter varied. Note also that only two visual elements are presented, and that these (similar to Shibuya & Bundesen's WR & PR study) are both followed by a mask.

From the results in [40] it seems that a SOA in the range from 0 to +500 ms is connected with a significantly lower identification rate for the secondly presented character. It is also worth noting that a SOA close to zero or in the order of 1000 ms seems to implicate that the identification rate of the second letter is not considerably lower than for the firstly presented character. In Figure 11 two characters were presented within each trial: a green digit (2 or 5) and a red letter (L or T). The characters were presented in an unpredictable order within a given trial, and the two characters

were never displayed in the same spatial position. Characters were presented for 45-60 ms (calibrated to the subject), and in average characters were presented for 57 ms (across all subjects). Each character was immediately followed by a mask lasting for 250 ms. For nine subjects total of 3 blocks of 156 trials were conducted, in total 4212 trials. Data shown in Figure 11 are averaged over all subjects, see p. 85 in [40].

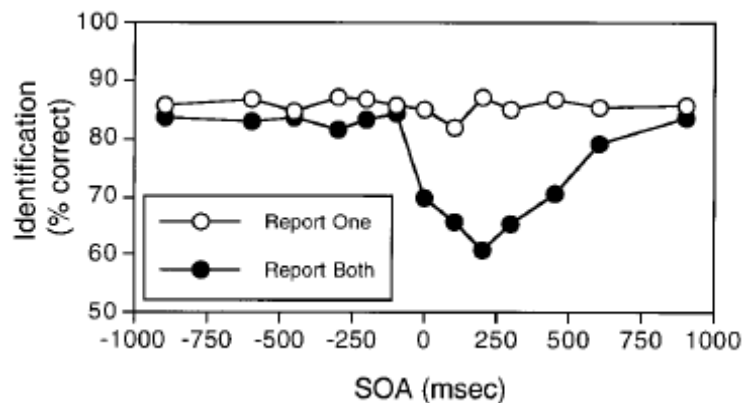


Figure 11: Accuracy of character identification as a function of SOA. Negative SOAs refer to accuracy on the first of the two characters presented; positive SOAs refer to accuracy on the character presented second. ‘Report Both’ conditions (filled circles): Both characters were identified. ‘Report One’ conditions (open circles): One character reported; other ignored. Figure adopted from [40].

From Figure 11 we see that when subjects are asked to report both of two serially presented characters, it seems that a difficulty arises to report the secondly presented character, i.e. the latter character is less likely to consolidate itself in VSTM.

In fact, and also from [40], we know that interference from earlier attended characters seem to increase with the number of attended objects. This latter phenomenon appears clearly in Figure 12 below. Here two or three characters were presented within each trial: one or two green digits (independently 2 or 5) and as before a single red letter (L or T). The characters were presented in an unpredictable order within a given trial, and no character was displayed in the same spatial position as any of the other characters. Characters were presented for 45-75 ms (calibrated to the subject), and in average characters were presented for 61 ms (across all subjects). Since methods were similar to Figure 11 we assume characters were immediately followed by a mask lasting for 250 ms. For 24 subjects a total of 4 blocks of 224 trials were conducted, in total 21504 trials. Again data shown are averaged over all subjects, see p. 93 in [40].

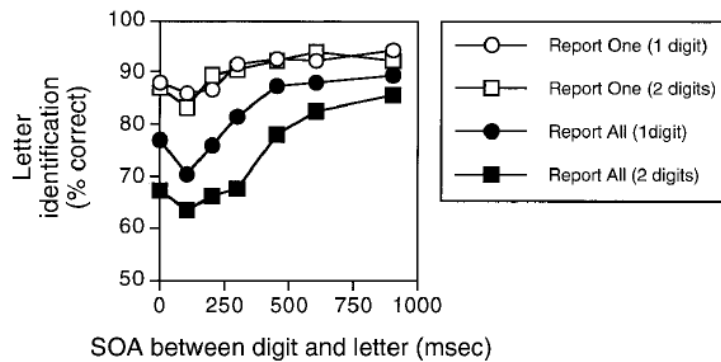


Figure 12: Letter identification accuracy as a function of SOA and the number of previously identified digits. ‘Report One’ (1 digit) (open circles): Ignore first character (digit), identify second (letter). ‘Report One’ (2 digits) (open squares): Ignore initial pair of digits, identify subsequent letter. ‘Report All’ (1 digit) (filled circles): Identify first (digit) and second (letter) characters. ‘Report All’ (2 digits) (filled squares): Identify initial pair of digits and subsequent letter. Note that only positive SOA is shown; this means that we only consider the identification of the secondly presented character – the letter. Figure adopted from [40].

As mentioned on p.95 in [40] it is striking that the interference from earlier presented characters does not seem to last considerably longer, but instead it increases in strength as more objects are initially attended. In relation to the important historical discussion about serial versus parallel processing in vision (cf. [26]), we think this suggests that the mechanism of attention can in this case be considered a parallel rather than serial process, where more objects share rather than line up for attention.

We shall note that the ‘Report All’ (1 digit) condition in Figure 12 is in principal similar to the ‘Report Both’ condition in Figure 11. Here we note however that in Figure 12 the identification rate on the latter character seems to be slightly less influenced from interference from the first character. This can be seen from the fact that the ‘Report All’ (1 digit) curve reaches a minimum of 70 %, whereas the ‘Report Both’ curve actually reaches a lower minimum of 60 %. Though Ward, Duncan and & Shapiro gives no direct explanation of this minor mismatch, we shall note that at least not all subjects were the same in the two experiments, and further in the latter experiment subjects reported vocally to the experimenter, while in the earlier experiment responses were made via a keyboard; a method of report that could perhaps be mentally more demanding to the subject.

PART II:
METHODS – REVIEWING VSTM MODELS

4 Fitting the storage capacity of VSTM in a whole report study

4.1 Fitting with the binomial distribution

We shall now inspect, how well the binomial distribution is able to model the typical score distribution depicted in Figure 4.

The binomial distribution can be characterized as a model that describes ‘sampling with replacement’. A typical example of this situation is, when a person is drawing balls from an urn that contains a number of black as well as a number of white balls. The distribution then denotes the probability that the person picks up a certain number of black balls in a fixed number of draws (see [2]). (Note that balls are put immediately back in the urn, after they have been drawn and that further it is an assumption of the binomial distribution that the number of balls has to be very large)

If the random variable X (the score) follows the binomial distribution $X \sim B(n, p_s)$, and we have n trials (storage slots), and the probability of a success (VSTM storage) is p_s , we can write the probability P of getting exactly j successes (a score of j) as:

$$P(j | n, p_s) = \binom{n}{j} p_s^j (1 - p_s)^{n-j} \quad \text{for } j = 0, 1, 2, \dots, n \quad (4.1)$$

In Kyllingsbæk’s whole report experiment scores varied from 1 to 5. However in the experiments by Shibuya & Bundesen (1988) we had scores of 6, and further some scientist have even reported a ‘magic number’ of 7, see [29]. Therefore we investigate three situations, where the number of storage slots n has either been fixed to 5, 6 or 7.

Because the binomial distribution is so fast to compute, we can simply compute X for many different parameter settings. Hence increasing p_s in discrete steps, we simply run through the interval $p_s \in [0; 1]$, and so for each step we compute the score X .

The fit that we end up with is the one that minimizes the summed squared error function SSE , which is calculated as:

$$SSE = \sum_{j=0}^n [P(j | n, p_s) - P_E(j)]^2 \quad (4.2)$$

Here $P_E(j)$ is the empirically observed probability for a score of j .

The binomial fits can be seen in section 11.1.

4.2 Fitting with the hypergeometric distribution

Next in this chapter we shall verify, whether the hypergeometric distribution is actually able to fit the data in Figure 4 better than the more often used binomial distribution.

The hypergeometric distribution is usually applied to model ‘sampling without replacement’. The classical example of this situation is, when a person is drawing balls from an urn that contains a number of black, as well as a number of white balls. The distribution then denotes the probability that the person picks up a certain number of black balls in a fixed number of draws, see [22].

In our case, if the random variable X (the score) follows the hypergeometric distribution $X \sim H(K, K_{tot}, n_{sa})$, and we have K successful scenarios (storage places), and we have K_{tot} scenarios in total (storage and non-storage places), and we make n_{sa} draws (storage attempts), then we can write the probability P of getting exactly j successes (a score of j) as:

$$P(j | K, K_{tot}, n_{sa}) = \frac{\binom{K}{j} \binom{K_{tot} - K}{n_{sa} - j}}{\binom{K_{tot}}{n_{sa}}} \quad \text{for } j = 0, 1, 2, \dots, n_{sa} \quad (4.3)$$

Similar to the previous section, we shall here investigate the three situations, where the number of storage places K has been fixed either to 5, 6 or 7. Also with the hypergeometric distribution we are dealing with a statistical distribution that is rather fast to compute, so again we perform an exhaustive calculation of the score distribution X as function of all parameter settings that we wish to explore. Our settings are:

$$K_{tot} \in 0, 1, 2, 3, \dots, K_{sim} \quad \text{and} \quad K \in 0, 1, \dots, K_{tot} \quad \text{and} \quad n_{sa} \in 0, 1, \dots, K_{tot}$$

As it appears, we limit our search for optimal parameter by setting $K_{sim} = 200$, which should be a sufficiently high value to find the optimum. The argument for this is that taken to the limit (as K_{sim} increases) we in practice obtain that n_{sa} becomes much smaller than $\min(K, K_{tot} - K)$. This means in fact that the hypergeometric distribution approaches the binomial distribution. In the classical example this corresponds to the case where the number of balls drawn from the urn is clearly smaller than both the number of black and white balls. Roughly speaking sampling with or without replacement is almost identical in large populations.

Again the fit that we end up with is the one that minimizes the SSE , calculated as:

$$SSE = \sum_{j=0}^{n_{sa}} [P(j | K, K_{tot}, n_{sa}) - P_E(j)]^2 \quad (4.4)$$

Here $P_E(j)$ is again the empirically observed probability for a score of j .

Also the hypergeometric fits can be seen in section 11.1.

5 Preliminary testing of Usher & Cohen's original model

5.1 The model

In section 1.2 we noted that in the NTVA interpretation, visual categorizations are thought to be encoded in VSTM by means of a positive feedback loop. In light of this notion a biologically plausible model for Short-Term Memory (STM) and selection was provided by Usher & Cohen in [37]. Usher & Cohen emphasize the difference between the anterior and the posterior part of the frontal lobe, and propose that while that memory in the first is re-activating itself, the memory in the latter is time-limited. Usher & Cohen's model focuses on the anterior part of the frontal lobe, which is thought to contain memory of more central and semantic nature, therefore they do not refer to their model specifically as a VSTM model, rather they simply call it a STM model

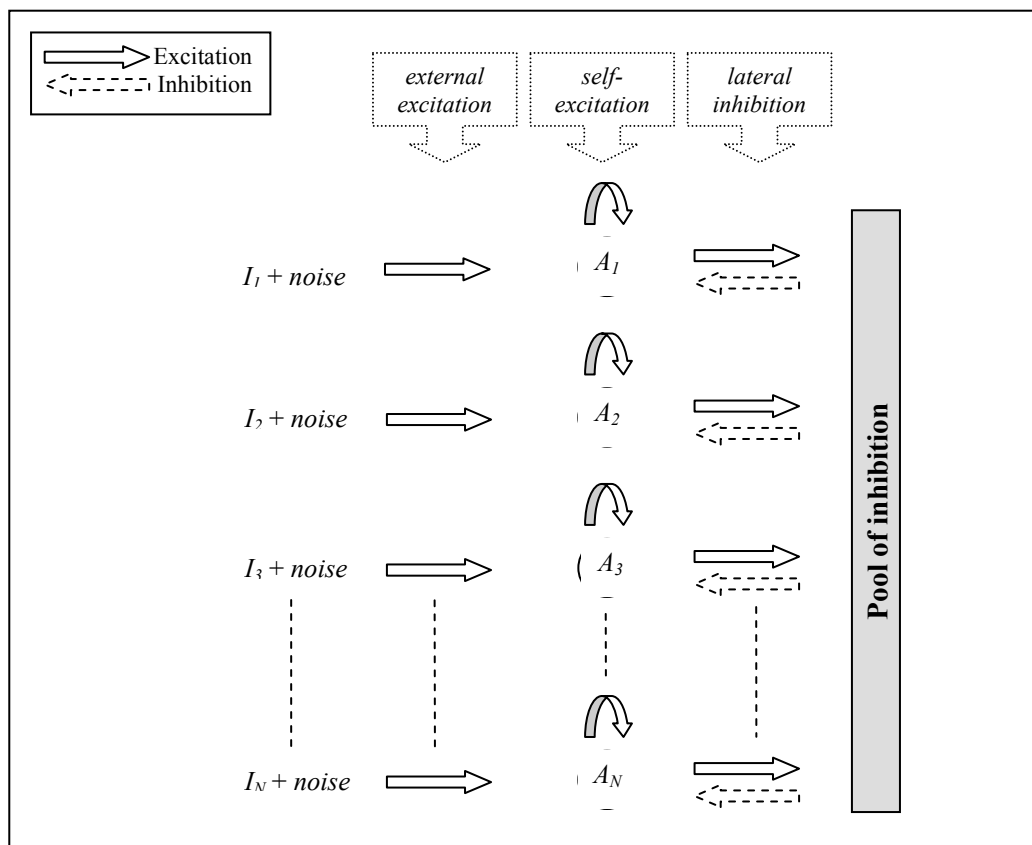


Figure 13: The principle in Usher & Cohen's STM-model. External excitation causes activation in the related assemblies (shown as circles). Once externally activated the assemblies sustain their activation due to self-excitation. Lateral inhibition keeps the general level of activation, seen over all assemblies, at a reasonably low level, i.e. effectively setting 'the size' of the memory.

Basically Usher & Cohen's model assumes that memory of an object can be interpreted as activation in an assembly of neurons. Each assembly of neurons corresponds to a unique object. If the activation of an assembly exceeds a certain activation threshold A_{th} memory of the object corresponding to that particular assembly is assumed.

The model in its basic form is in fact rather simple, which means that it can actually be contained in one line. Consider the activation A_x of the cell assembly x :

$$\frac{dA_x}{dt} = -A_x + \alpha^* F(A_x) - \beta^* \sum_{z \neq x} F(A_z) + I_x + noise \quad (5.1)$$

The above equation characterizes a leaky accumulator model. There is passive decay of the activation towards the rest level, with a time constant chosen as 1, reflecting the time scale that physiologically is observed with synaptic currents [37].

Here α^* is a constant scaling the self-excitation of the assembly, and β^* is a constant scaling the lateral inhibition coming from all other assemblies than itself. I_x is the external input that supports activation in the assembly x and *noise* is some kind of external noise source, presumably Gaussian white noise. $F(A)$ is an activation function that keeps the activation within bounds, i.e. $F(A) \in [0;1[$.

A sigmoidal is chosen for the activation function, and it is defined as follows:

$$F(A) = \begin{cases} 0, & \text{for } A \leq 0 \\ \frac{A}{1+A}, & \text{for } A > 0 \end{cases} \quad (5.2)$$

One could notice that the values of α^* and β^* do not depend on the specific assembly considered; rather these constants are assumed the same throughout the entire network.

In previous use of Usher & Cohen's model the external input I_x has generally been kept at a constant level higher than zero, in what one could think of as the active period, while in the rest of the simulation the input is kept at zero.

5.2 Implementation

Usher & Cohen used Euler integration with an integration constant $\lambda = 0.99$, hence in the discrete time step s , their update formula can be written as:

$$A_x(s+1) = \lambda(A_x(s)) + (\lambda - 1) \left(\alpha^* F(A_x(s)) - \beta^* \sum_{z \neq x} F(A_z(s)) + I_x(s) + noise(s) \right) \quad (5.3)$$

To illustrate the behavior of Usher & Cohen's STM-model we implemented the model in Simulink (see description in [37]). The model was tested with settings corresponding to parallel as well as serial presentation of input. All inputs were similar in both amplitude I and in duration (the number of time steps the input was kept at I). In Figure

14 we see a situation with parallel presentation of inputs. Here three inputs manage to consolidate themselves, i.e. they end up with a positive activation higher than the activation threshold $A_{th} = 0.2$. As all inputs are identical the winners can only be selected because additive noise was included in the model. In Figure 15 the simulation settings were identical to Figure 14, but rather than three now only two inputs manage to consolidate themselves.

Figure 16 and Figure 17 are very similar to Figure 14 and Figure 15. The only difference in the settings of the parameters is that the level of inhibition has been lowered from $\beta^* = 0.2$ to $\beta^* = 0.1$. We see that it now takes more time to select the winners, but still the number of winners vary for the same reason as for the previous plots. Due to reduced competition in the network we now see more inputs consolidating themselves.

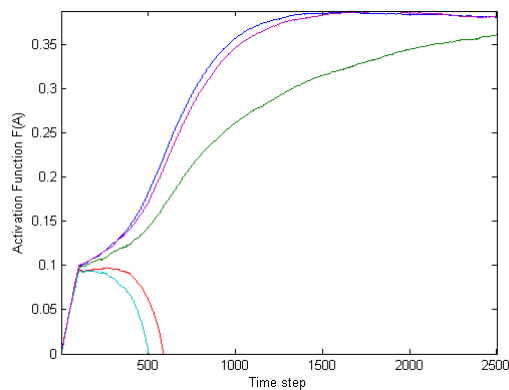


Figure 14: Parallel presentation of inputs with $\beta^* = 0.2$. Reproduction of upper plot in Figure 3 in [37]. Input in active period was clamped to $I = 0.01$.

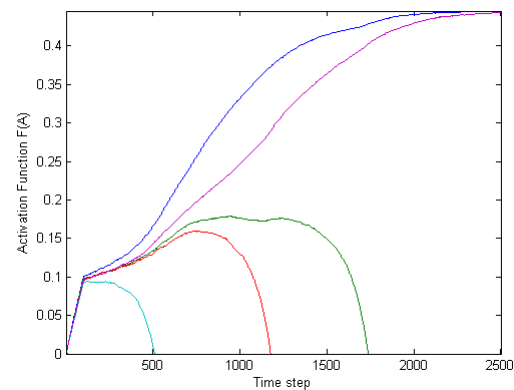


Figure 15: Again parallel presentation of inputs with $\beta^* = 0.2$. Reproduction of upper plot in Figure 3 in [37]. Input in active period was clamped to $I = 0.01$.

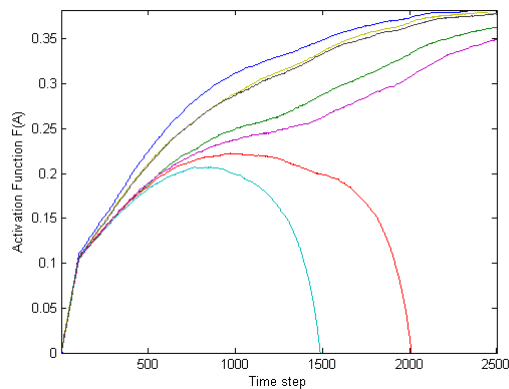


Figure 16: Parallel presentation of inputs with $\beta^* = 0.1$. Reproduction of lower plot in Figure 3 in [37]. Input in active period was clamped to $I = 0.01$.

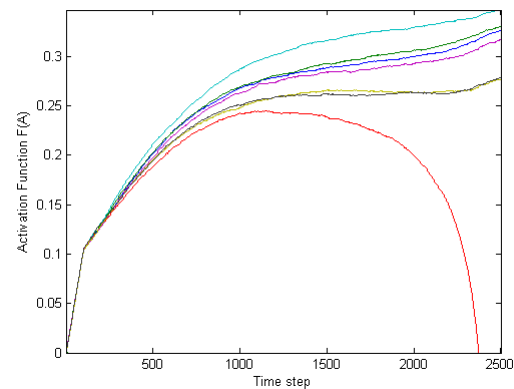


Figure 17: Again parallel presentation of inputs with $\beta^* = 0.1$. Reproduction of lower plot in Figure 3 in [37]. Input in active period was clamped to $I = 0.01$.

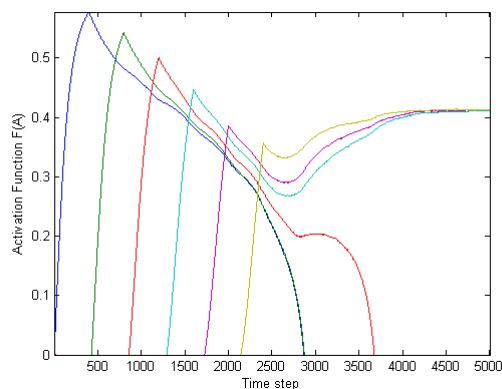


Figure 18: Serial presentation of input.
Reproduction of plot in Figure 4 in [37].

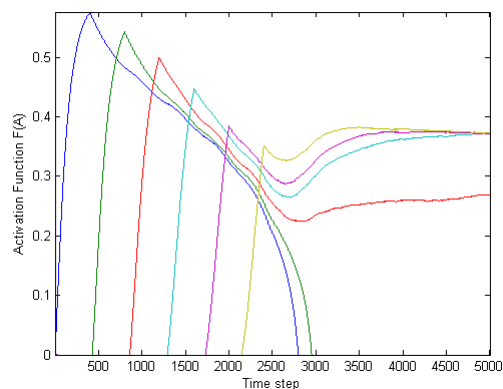


Figure 19: Serial presentation of input.
Reproduction of plot in Figure 4 in [37].

In Figure 18 and Figure 19 Usher & Cohen's model was tested with a serial presentation situation. Inputs were presented for 101 time steps, and the onset interval between the inputs was fixed at 400 time steps. From the plots we see that the firstly presented input rises to the highest activation level. This is due to the fact that no lateral inhibition is received from other input patterns. As more inputs are presented their immediate rise in activation level is more and more limited, due to the fact that more and more lateral inhibition is received from previously presented inputs. At a certain time the winners are selected and the losers die away. We see again that the number of winners vary, but with the current settings 3 to 4 winners is not uncommon. Also we see that generally it is the lastly presented inputs that succeed in consolidating themselves. This is consistent with experimental studies, and further it is a consequence of the setting of the parameters.

All plots in Figure 14 to Figure 19 were reproductions of plots shown in [37]. Encouragingly our plots closely resemble Usher & Cohen's original plots.

5.3 Critique of Usher & Cohen's model

It does not seem very plausible that there exist only a small number of possible objects N_{na} that can become encoded into VSTM. From every-day life it is our experience that an infinite number of different objects can in fact occupy our visual attention and become encoded in VSTM, though not at the same time!

However, considering Usher & Cohen's model the number of neuron assemblies, N_{na} included in the model ought not to be of great practical importance. This claim is due to the fact that neuron assemblies that are not activated by an input are kept in the inactive state, where they do not feed any lateral inhibition to other neuron assemblies, i.e. in the model deactivated neurons do not interfere with other neuron assemblies. The neuron assemblies that are not activated by regular input can however still be activated by the random white noise that was added. As we understand the model, the number of assemblies N_{na} must have at least a minor importance in Usher & Cohen's original model! This possibly means that maximal level of self-excitation α^* , and the maximal

level of inhibition β^* can not be set independently of the number of assemblies, which is in our opinion highly impractical!

In order to investigate the effect of the setting of N_{na} , a number of series of simulations are conducted; each with a fixed value of N_{na} . Results from this preliminary investigation of Usher and Cohen's model are shown in section 11.2.

We activate all N_{na} assemblies initially. The assemblies receive an input $I = 0.1$ for the first 101 time steps. Every time step has a size of 0.01, and the simulation stop time is set to 25.

Further we apply standard normal distributed random noise scaled with a constant of 0.02. This noise is added to the input at all times. We use standard Gaussian white noise. The noise is not in any way dependent upon the level of the input.

The activation threshold A_{th} is set to 0.2, so that all objects that end up with having a level of activation above this threshold are considered as the ones stored in STM. (We consider the activation in the last time step of the simulation.) For every setting of the parameters α^* and β^* we calculate the mean number of stored inputs. (The mean value is calculated based on a total of 40 repetitions per setting.) The results from this investigation are shown in Figure 38 and in Figure 39.

Further a point of critique for Usher & Cohen's model is that the activation threshold is set more or less in an arbitrary fashion. In the spike model we shall later present we actually avoid this rather peculiar problem.

6 The two-component mixture model (2CMM and 2CMM₂)

6.1 Shibuya & Bundesen's original version

After considering related classes of models as Independent Parallel Models and Independent Serial Models, Shibuya & Bundesen concluded that a Fixed-capacity Independent Race Model (FIRM) (cf. [6]) was suitable for fitting the WR & PR data reported in [38]. The model explains the data on the hypothesis that efficiency of visual selection (the efficiency of selecting targets rather than distractors) is independent of exposure duration.

The equations for FIRM are reported in Appendix A. First of all we shall note that FIRM is a model that is built purely upon straightforward probability theory. The equations in Appendix A are based on the assumption of homogenous objects, which means that all targets have the same attentional weight w_t and all distractors have the same attentional weight w_d . In section 2.5 we saw that the processing capacity C can be treated as a constant, hence the rate of encoding for targets, v_T and for distractors, v_D can be written as:

$$v_T = \frac{C}{T + \alpha D} \quad v_D = \frac{\alpha C}{T + \alpha D} = \alpha v_T \quad (6.1)$$

Where T is the number of targets in the display, and D is the number distractors in the display. The rate of discrimination α characterizes how capable the subject is at discriminating between targets and distractors.

In FIRM one assumes a fixed capacity K , which could be either 3 or 4 objects. Shibuya & Bundesen however implemented a version of FIRM capable of handling non-integer numbers for K . (See [38]). Here for instance a value of 3.74 would be treated as a mixture of values 3 and 4, with a probability of 0.74 for having K set at 4. Thus by implementing a simple 2-component mixture model, Shibuya & Bundesen improved the resolution in the FIRM without adding extra parameters.

FIRM assumes that the effective exposure duration τ is smaller than the actual exposure duration t_e by an amount t_0 corresponding to the temporal threshold, before conscious processing begins. However the effective exposure duration can not be negative so computationally it is set to:

$$\tau = \max(0, t_e - t_0) \quad (6.2)$$

6.2 Optimization method

We apply constrained nonlinear optimization in order to determine the optimal mixing coefficients for the two-component mixture model. (More information can be found in the documentation for the standard Matlab function *fmincon*.)

We define an experimental condition $\theta \equiv \{T, D, \tau\}$ as characterized by the number of targets T , the number of distractors D and the exposure duration τ .

Since we are dealing with a mixture model, the *model* conditional probability $p_m(j|\theta)$ that the score is equal to j (given θ) equals the linear combination of component densities $p_m(j|\theta, K)$ weighted with their respective mixing coefficients $p(K)$:

$$p_m(j|\theta) = \sum_K p_m(j|\theta, K)p(K) \quad (6.3)$$

Hence the likelihood of the dataset L can be written as:

$$L = \prod_{\theta} \prod_j (p_m^*(j|\theta))^{n_e(j|\theta)} \quad (6.4)$$

Where $n_e(j|\theta)$ is the number of *empirical* observations of the score j (given the experimental condition θ).

The cost function E that we minimize is the negative log-likelihood, which is written as:

$$E = -\log(L) = -\log\left(\prod_{\theta} \prod_j (p_m^*(j|\theta))^{n_e(j|\theta)}\right) = -\sum_{\theta} \sum_j n_e(j|\theta) \log(p_m^*(j|\theta)) \quad (6.5)$$

In order not to risk taking the logarithm of zero we verify that $p_m(j|\theta)$ is not equal to zero. If this is the case $p_m(j|\theta)$ is simply reset to 1, implying that the term cancels itself out in the cost function.

The initial setting of the free parameters that we apply is:

$$C = 50 \text{ s}^{-1} \quad \alpha = 0.4 \quad t_0 = 0.020 \text{ s} \quad P(K) = 0.5$$

Where we note that: $K \in \{3,4\}$.

The constraint we use for the optimization is:

$$\sum_K P(K) = 1 \quad (6.6)$$

Further we demand that:

$$C \in [1s^{-1}; 500s^{-1}] \quad \alpha \in [0.1; 1] \quad t_0 \in [0.015s; 0.025s] \quad P(K) \in [10^{-4}; 1]$$

Here we shall note that the setting of the initial values as well as the defined boundaries for C , α and t_0 are directly inspired by the optimal values found by Shibuya & Bundesen in [38]. Further for computational causes, i.e. in order not to have terms in the cost function vanish, we demand that $P(j) > 0$.

The maximum number of iterations per optimization is set to 1000. First the optimization is started with the above stated initial values. After this the algorithm is restarted 3 times in the most recently found minimum. It is the result from the third restart that we consider as the final optimization result.

6.3 Critique of Shibuya & Bundesen's optimization method

Similar to us Shibuya & Bundesen used an iterative method for searching the space of parameters. Their optimum fit was found also by optimizing a maximum likelihood cost function. One of the discrepancies they reported was that their model was simply not able to re-produce scores of 5 or more, given that K is set below or equal to 4. In [38], Subject MP however scored 5 on a total of 6 trials, and hence Shibuya & Bundesen made two computational assumptions:

- 1) One of the six trials was under the experimental condition $T = 6$, $D = 4$ and $\tau = 200$ ms. Shibuya & Bundesen chose to neglect this single trial, because the relative frequency of the score 5 was so small (i.e. 0.0167).
- 2) The other 5 trials, where a score of 5 was reported, happened under the experimental condition $T = 6$, $D = 0$ and $\tau = 200$ ms. Here Shibuya and Bundesen could no longer neglect scores of 5 due to infrequency (the relative frequency was 0.0833). Instead they chose to proceed as if the trials with a score of 5 could actually be counted among the trials, where the score had only been 4.

Strictly speaking judging whether a model is likely or not (i.e. considering ML), a model has to be able to reproduce any empirical observation (at least with some small probability), if not the model should simply be rejected.

One could argue that instead of having both of the above two mentioned computational assumptions one should be consistent and for example count all observations, where a score above 4 was reported, as if the subject had scored 4. For comparison we made a corrected fit to the empirical data with Shibuya and Bundesen's 2-component mixture model, now simply counting all trials where scores above 5 were observed, as if the score had been 4. Without the just mentioned correction we denote the model: the two-component mixture model (2CMM), and with the correction we denote it: the corrected two-component mixture model (2CMM₂)

For comparison Figure 20 shows the fit obtained by Shibuya & Bundesen together with the corrected fit.

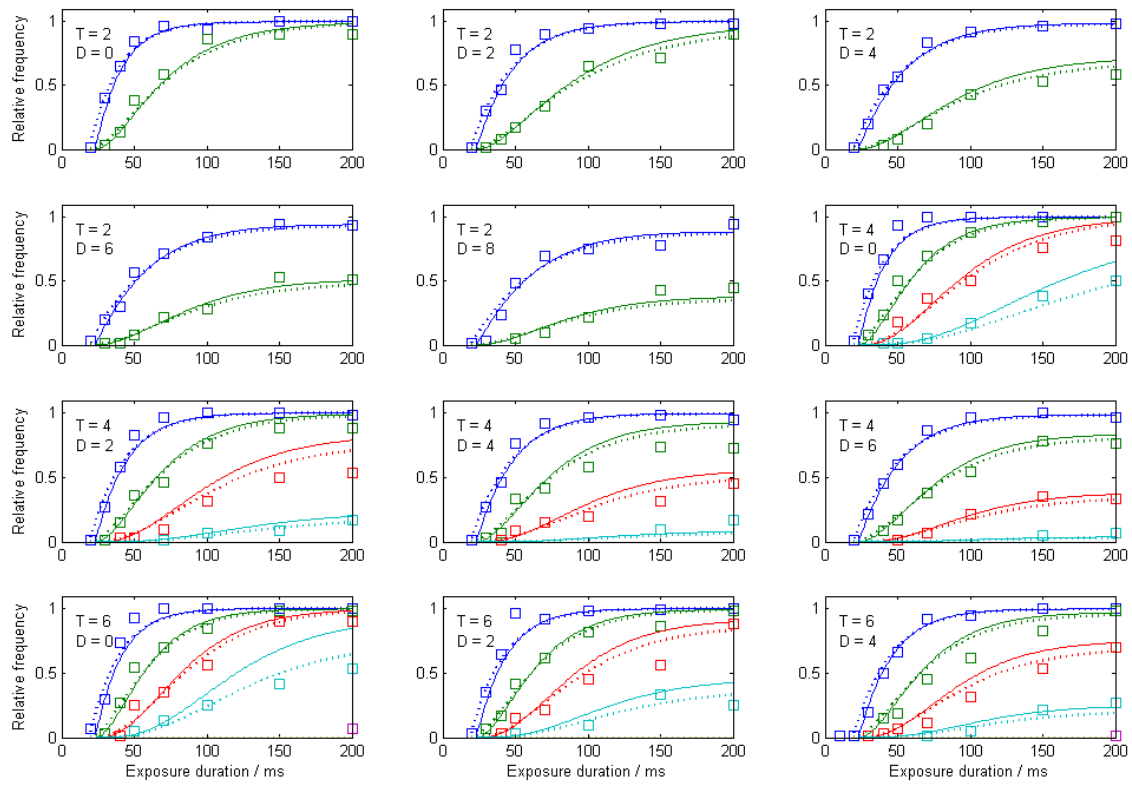


Figure 20: Dotted lines represent Shibuya & Bundesen's original fit with the two-component mixture model. Solid lines represent the corrected fit with the same model now counting all scores of 5, as if the score had been 4.

7 The multi-component mixture model (MCMM)

7.1 An improved mixture model

It is clear that a better fit can sometimes be produced by expanding a mixture model to include more components.

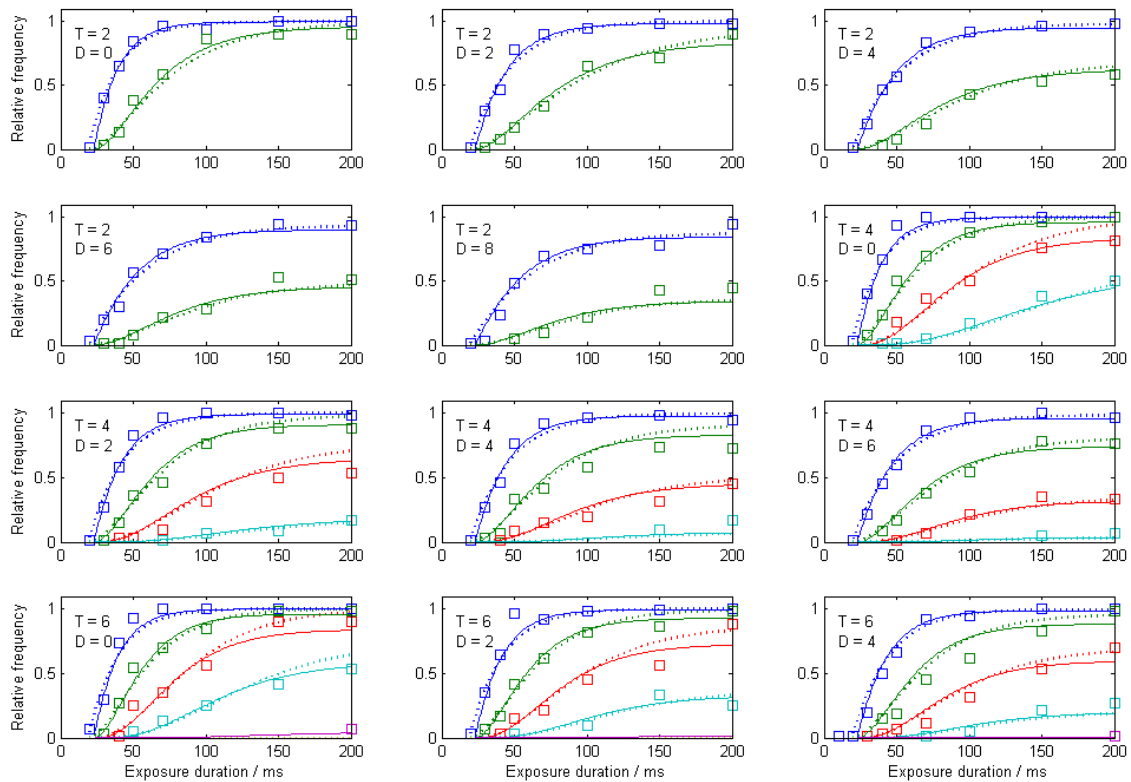


Figure 21: Dotted lines represent Shibuya & Bundesen's original fit with the two-component mixture model. Solid lines represent the fit with the 5-component mixture model.

In Figure 4 we expanded Shibuya & Bundesen's model to cover components K of 1, 2, 3, 4 and 5. Hence also this model is based on the FIRM equations found in Appendix A. It is seen from a qualitative point-of-view that this 5-component mixture model generally fits the empirical data better than Shibuya & Bundesen's original 2-component mixture model.

7.2 Optimization method

Mixing coefficients for the multi-component mixture model was determined with the same optimization technique as described above for the two-component mixture model.

Now however we naturally have that: $K \in \{1,2,3,4,5\}$, and further we make the initial setting: $P(K) = 0.2$ in order to insure that we also initially fulfil the constraint given in equation (6.6).

8 The unit spike model (USM)

8.1 Modifying Usher & Cohen's model

The model described in this chapter is directly inspired by the STM model presented by Usher & Cohen in [37].

Considering Usher & Cohen's model for STM (See equations (5.1) and (5.2)), we noted that input I was always kept at a constant level. The level was set to a constant higher than zero, if the input was present. If input was absent the level was simply set to zero. Suppose all inputs were given the same attentional weight and further suppose the inputs were presented in exactly the same time interval, then all input levels would be equal at all times. In this situation it is absolutely essential that Usher & Cohen include a term representing additive noise. The noise puts a unique fingerprint on each input effectively helping a subset of the otherwise equipotent input candidates to win the race.

We shall now form the hypothesis that the input to the competitive leaky integrator is not simply a constant as explored by Usher & Cohen in [37]. Rather in our model we propose that the input takes the form of Poisson distributed spike trains. Following Bundesen (1990) we interpret the encoding rate v_x for the object x as the rate of a Poisson spike generator G . This now leads us to a slightly modified version of equation (5.1):

$$\frac{dA_x}{dt} = -A_x + \alpha * F(A_x) - \beta * \sum_{z \neq x} F(A_z) + G(v_x) \quad (8.1)$$

We see that another change, which we have made, is that the additive noise has been excluded (see Figure 22). The reason for removing the noise term is justified by the fact that Poisson processes themselves induce a sort of noise, from which a subset of equipotent input candidates can be selected.

Further in our modified model we simply set the threshold A_{th} that the activation A_x has to cross before the object x is considered to be stored in VSTM to zero. In contrast, in Usher & Cohen's implementation the threshold was more or less arbitrarily set to some small positive constant, e.g. 0.2. Though Usher & Cohen [37] claim that the setting of the activation threshold does not influence their results, we speculate that the reason for choosing a value like 0.2 was simply to have as low a threshold as possible, still ensuring that: $A_{th} < |noise|$.

If the activation threshold was not kept at a higher level than the noise amplitude, then the noise alone would be sufficient to encode completely random object into memory. Hence encoding could take place even without any input, i.e. without any visual elements being present in the display!

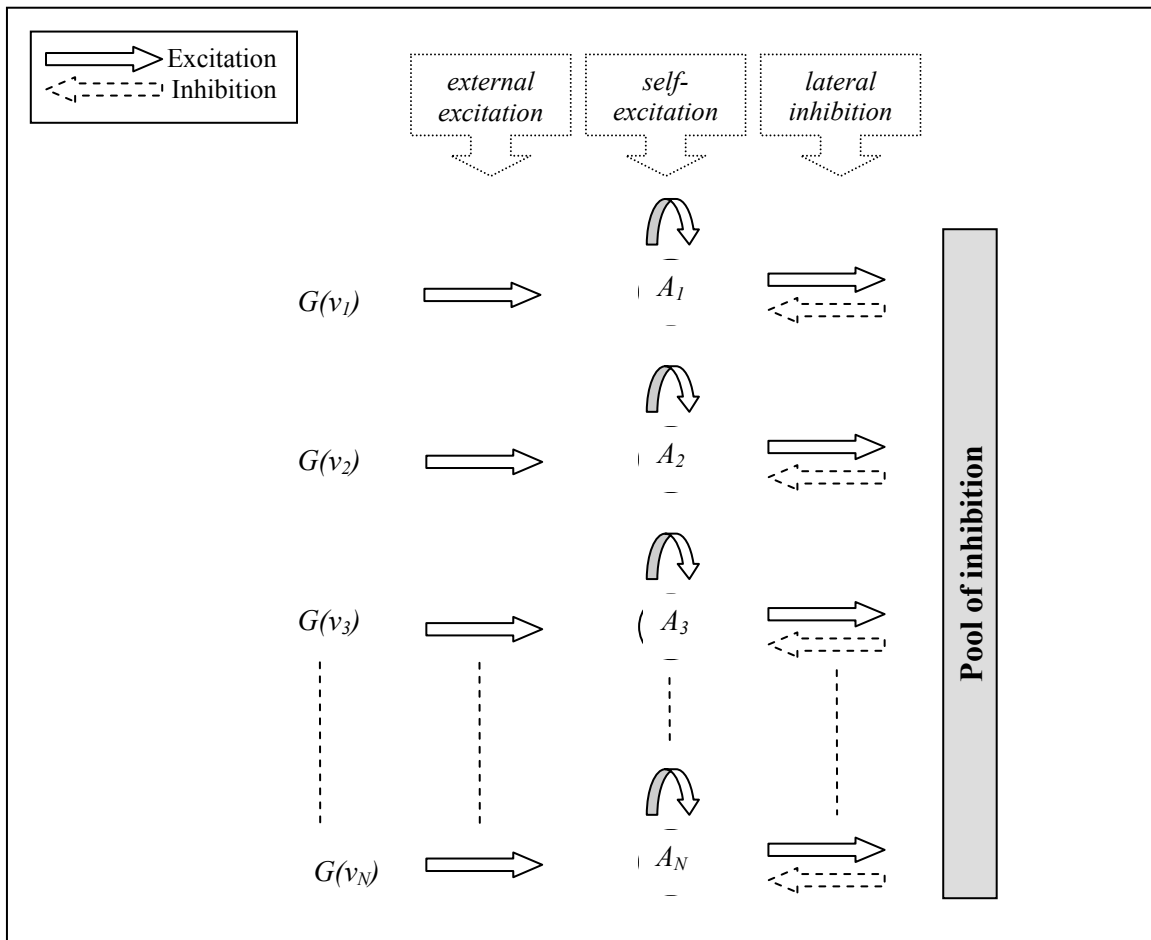


Figure 22: The principle of the Unit Spike VSTM model. The model is inspired by Usher & Cohen’s STM model (see Figure 13). However the inputs have been substituted with Poisson generators G spiking with rate parameters v . Further additive noise is no longer included and the activation threshold A_{th} has been set to zero.

We note that we do however not have any problem with the activation threshold in our modified model. Here it is only the activated objects (i.e. the objects that are actually present in the visual field) that receive a noise input, non-activated objects representations do not receive any input at all, not even noise. Hence it makes sense to keep the model as straightforward as possible, and therefore we simply set the activation threshold A_{th} to zero.

8.2 Adopting the parameters: C , t_0 and α from Shibuya & Bundesen

In our model we adopt the parameters C , t_0 and α from the previously described mixture models. In order to demonstrate that this is a feasible way to go, we conduct one single simulation with the multi-component mixture model and seven simulations with the unit-spike VSTM model. In Figure 23 dotted curves (MCM) are identical in all the plots, while the solid curves (USM) are not.

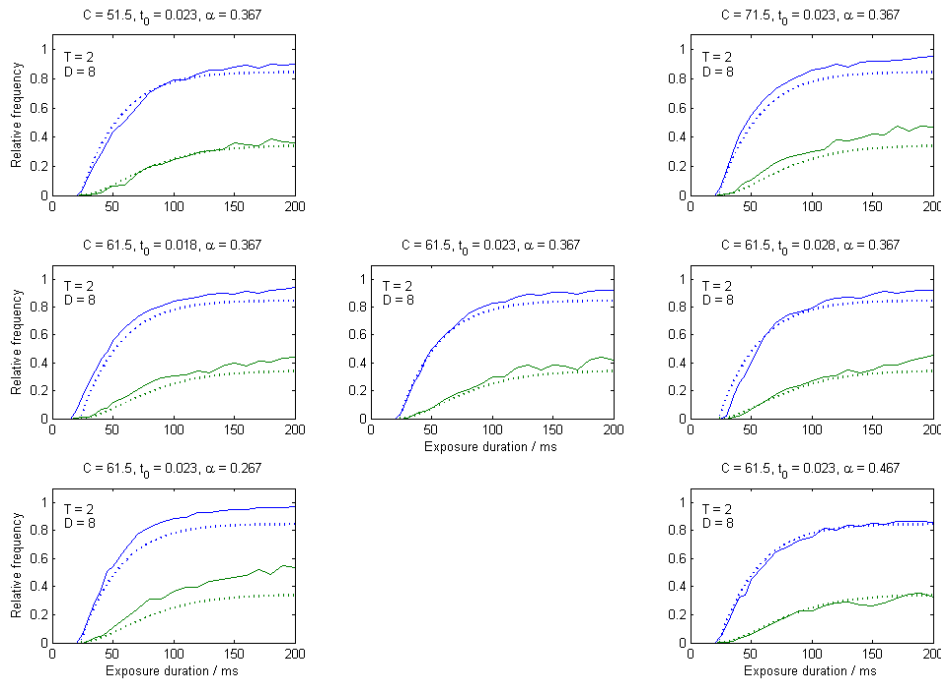


Figure 23: Cumulated score diagrams. Dotted lines represent a fit with the multi-component mixture model. Solid lines represent a fit with the unit-spike model. The plot placed in the middle corresponds to the situation where C , t_0 and α match the values we obtained by fitting the multi-component mixture model. In the top row the parameter varied is the processing capacity C . In the middle row the parameter varied is maximal ineffective exposure duration t_0 . Finally in the bottom row the parameter varied is the ratio of discrimination α .

The reason that we have correspondence between the parameters of the two different types of models is that for small effective exposure durations τ , the spike model behaves actually as the mixture model. In both type of models the probability that a visual element (i.e. a target or a distractor) enters VSTM equals the chance of a random point event drawn from a Poisson process with a rate parameter ν . (For the mixture model it is here assumed that the memory is not already occupied by other encoded elements.) The reason that the spike model offers this behaviour is that for small τ we can assume that externally activated objects have so far only received negligible self-excitation. Further and as a consequence of this the amount of lateral inhibition is thereby also negligible. Hence for small τ we can simply assume that the level of activation is zero for all objects, i.e. one single spike is enough to encode an object in VSTM.

Though the spike model is different than the mixture model the parameters: C , t_0 and α cover similar concepts in the two types of models. We propose that C , t_0 and α are generally determined with the mixture model. The parameters can then be directly transferred or alternatively used as an initial guess for the same parameters in the spike model. This procedure makes sense, since it is much faster to optimize parameters in the mixture model than it is in the spike models.

Also we should note that, following Shibuya & Bundesen's procedure for the mixture model, also for the spike model we make use of equations (6.1) and (6.2) in order to determine rates of processing ν and the effective exposure duration τ .

8.3 Validity of simulation and model parameters

It is an inherent limitation of our implementation that there can only be 1 spike per time step. Consider the extreme case, where all processing capacity C is occupied by one single object, i.e. $v = C$. In this case, the probability of having n spikes arriving in a time interval Δt is then given by:

$$P_n = (v \cdot \Delta t)^n \quad (8.2)$$

This means that with a processing rate of 60 s^{-1} the chance that we have more than one spike in a time interval of 1 ms is equal to:

$$P_2 = (v \cdot \Delta t)^2 = \left(60 \frac{1}{s} \cdot 1 \text{ ms}\right)^2 = 0.36\% \quad (8.3)$$

Based on the small probability for having more than 1 spikes we consider the step size of 1 ms to be sufficiently small (at least for processing rates up to 100 s^{-1} , where $P_2 = 1.00\%$).

8.4 Shibuya & Bundesen's data – Estimating parameters for the model

The data set of Shibuya and Bundesen is in fact rather large. They used a total of 12 different combinations of T and D, and further for each combination they used 9 different exposure durations. This leaves us with a total of 108 different experimental settings. The fact that the VSTM seems to have a stochastic rather than a deterministic nature means that for each setting one must have many trials. Hence for each of the 108 settings 60 trials were conducted. Thus the data for subject MP consists of a total of 6480 trials.

We do not know any explicit function expression of the model, so for every set of model parameters we want to test; we simply run to model a number of times in order to evaluate, how the model responds. Due to the stochastic nature of the model we also need to have many trials per setting in order to get a reasonable estimate, of how the model performs.

In its current implementation the simulation takes about 0.2 seconds in physical terms (i.e. the time used by the computer). We used a simulation stop time of 2.5 seconds (which is of course in model terms!).

In order to determine optimal parameters for the USM we fitted the model to the WR and PR data set of subject MP in [38]. For this we used unconstrained nonlinear optimization as offered by the standard Matlab function *fminsearch*. The parameter *MaxIter* was fixed to 20. This function effectively implements the useful Nelder-Mead Simplex algorithm [27]. Depending on the initial condition that we used with the

algorithm, we obtained very different parameter sets in the optimization. An explanation for the varying results that we obtained is presented in Figure 24, which shows also an error map. The algorithm was restarted 10 times, indicated by the red squares in the figure. For each parameter setting the algorithm was run 200 times, before the mean distribution was calculated. We use the logarithm of the summed squared error as our error cost function, since it seems to yield the error map with the nicest contour lines. The shown error function was chosen over the summed squared error and the negative log likelihood error functions. In the plot the error function has not been smoothed with a mean filter, though this could be a way to remove noise caused by the Poisson distributed input processes.

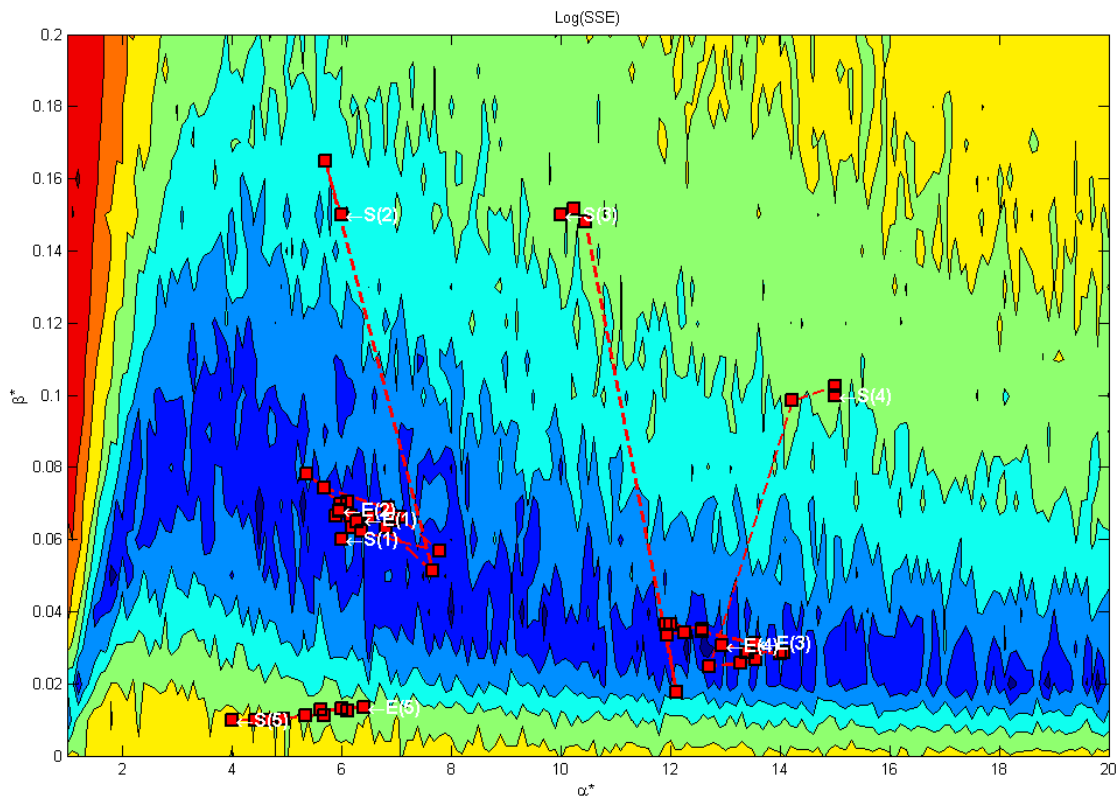


Figure 24: Error map of the Unit Spike Model. Dark blue areas are the ones with the smallest associated error. Five optimization runs were conducted, the starting point for these were S(1) to S(5) and the ending points E(1) to E(5), correspondingly. We see that all optimization runs approach the error optimal (dark blue) area. The fifth run however, only reaches a green region, indicating that the algorithm has not been restarted enough times in order to converge. A total of 200 trials were conducted for each setting of the model parameters in order to compute the map.

As it is seen from Figure 24, there is a valley in the error map. This indicates that for any $\alpha^* \in [0; 20]$ there seems to be a value $\beta^*(\alpha^*) \in [0, 0.2]$, so that the error function is close to the global minimum (here we are speaking of the global minimum for the plotted parameter domain, i.e. for the domain where $\alpha^* \in [0; 20]$ and $\beta^* \in [0, 0.2]$).

The error function was calculated as:

$$\log(SSE) = \log \left(\sum_{K=1}^5 \sum_{\theta} \left(\overline{P}_M(K | \theta) - P_E(K | \theta) \right)^2 \right) \quad (8.4)$$

Here the experimental conditions that we consider are all empirically tested conditions $\theta = \{T, D, \tau\}$ from [38], where the exposure duration τ was fixed at 200 ms. The reason that we only included the latter part of the experimental conditions was that otherwise the simulation time would be increased to about two weeks (i.e. an increase with a factor of 9).

8.5 Dynamics of the model

Concerning the dynamics of the model it is important first to note that any visual element (target T or distractor D) that ends up with a positive activation level A (i.e. above the activation threshold A_{th}) is considered as stored in visual short-term memory. In Figure 25 below we see activation levels depicted as a function of time for four identical simulations. The model effectively implements a sort of winners-take-all neural network.

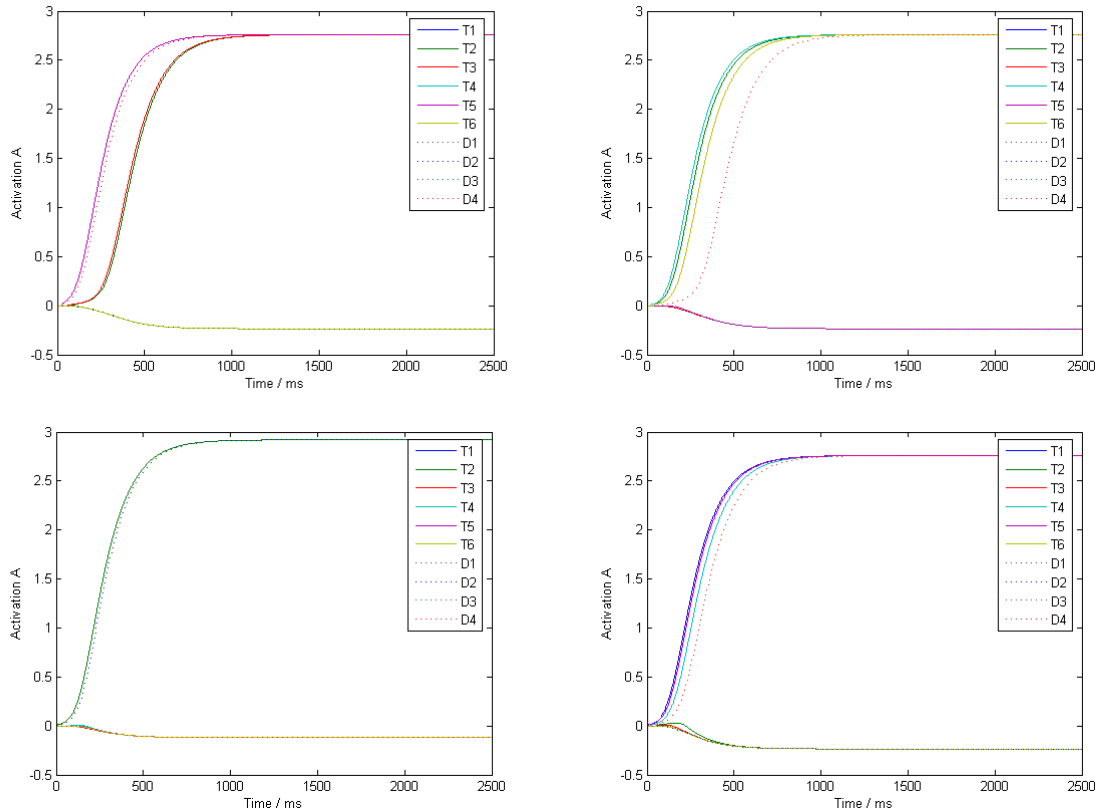


Figure 25: Dynamics of the Unit Spike Model. Parameters are the ones mentioned in Table 2 for the USM. The four plots represent activation levels as a function of time. Experimental conditions for all four plots were: $T = 6$, $D = 4$ and $t_e = 200$ ms. The four simulations plotted were selected at random.

From the plots in Figure 25 we note that in the first 200 ms visual elements try to encode themselves into VSTM with the help from the Poisson distributed spike trains. When the first visual element has fired a spike, the activation level of the element will start to increase due to self-excitation of the cell assembly. As more and more element fires a spike and manages to increase their activation level, the lateral inhibition in the network rises. This means that elements that do not fire a spike are suppressed by the high activation of the rest of the elements. Also elements that only manage to spike one time are often vulnerable to lateral inhibition from other activated elements. This latter observation is especially true, if the single fired spike is placed in the last part of the exposure duration. Note also that targets are more likely to be encoded in VSTM. This is due to the influence from the discrimination factor α that implements the fact that there is a lower probability for firing a spike for distractors than for targets.

9 The non-unit spike model (NUSM)

9.1 *Modifying the unit spike VSTM model*

From Usher & Cohen (1999) we know that there is passive decay of the activation towards the rest level. The time constant for the decay is chosen as 1, reflecting the time scale that physiologically is observed with synaptic currents [37]. Hence we do not feel any need to include a scaling factor for this term.

On the other hand as noticed from (8.1), we do have scaling factors for both the self-excitation and lateral inhibition, the scaling factors being α^* and β^* .

As the model appeared in (8.1), we implicitly assumed that the amplitude of the input spikes was appropriately set in comparison to the time constant for the passive decay. Our message in this chapter is that we might obtain a better model, if we include a scaling factor γ^* for the input spikes.

Simply expanding the model described in the previous chapter we obtain a modified model described by

$$\frac{dA_x}{dt} = -A_x + \alpha^* F(A_x) - \beta^* \sum_{z \neq x} F(A_z) + \gamma^* G(v_x) \quad (9.1)$$

9.2 *Shibuya & Bundesen's data – Estimating parameters for the model*

For reasons of computational burden we did not proceed to optimize the parameters of the non-unit spike model using a standard non-linear optimizer. Instead our choice was again to plot an error map for the model in order to get a better understanding of, where the optimal parameter regions were located. The plots in Figure 26 below were all produced in a similar way as described in section 8.4 above. Thus the error function that we used was again the one presented in equation (8.4). In Figure 26 below we fixed the spike amplitude γ^* at 6 different values in order to get an understanding of what is actually (in an error point-of-view) the significance of this parameter.

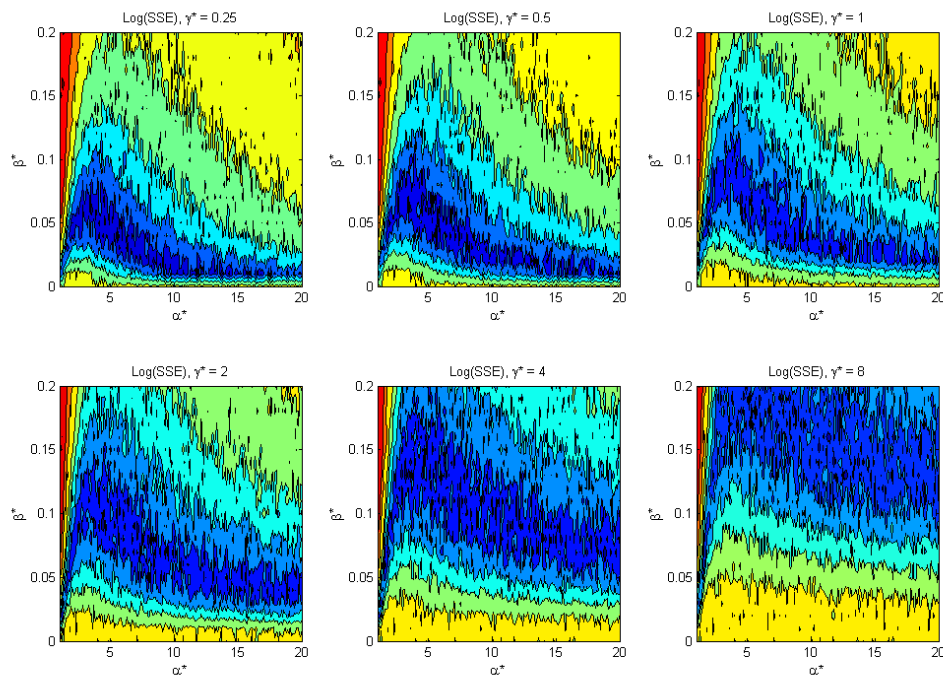


Figure 26: Cost function dependency as a function of α^* and β^* . Each window shows the cost function for a fixed value of $\gamma^* \in [0.25; 8]$. The dark blue areas represent the optimal parameter regions from an error point-of-view. Notice that the upper right plot is of course similar to the plot in Figure 24. A total of 200 trials were conducted for each setting of the model parameters.

As we see from Figure 26, there is some evolution in how the error function appears, as we increase the spike amplitude γ^* . All plots are very similar to the plot that we saw for the USM (the upper right plot). Generally we need to increase the maximal level of inhibition β^* , when we increase the spike amplitude γ^* . Also we can note if γ^* is increased the error optimal region (conditioned on the fixed parameter γ^*) seems to increase in size.

9.3 Dynamics of the model

Concerning the dynamics of the model everything that was mentioned in section 8.5 for the USM also applies to the NUSM. In Figure 27 below we see again activation levels depicted as a function of time for four identical simulations. The model effectively implements a sort of winners-take-all neural network. Experimental conditions were similar to the ones used in section 8.5 for USM.

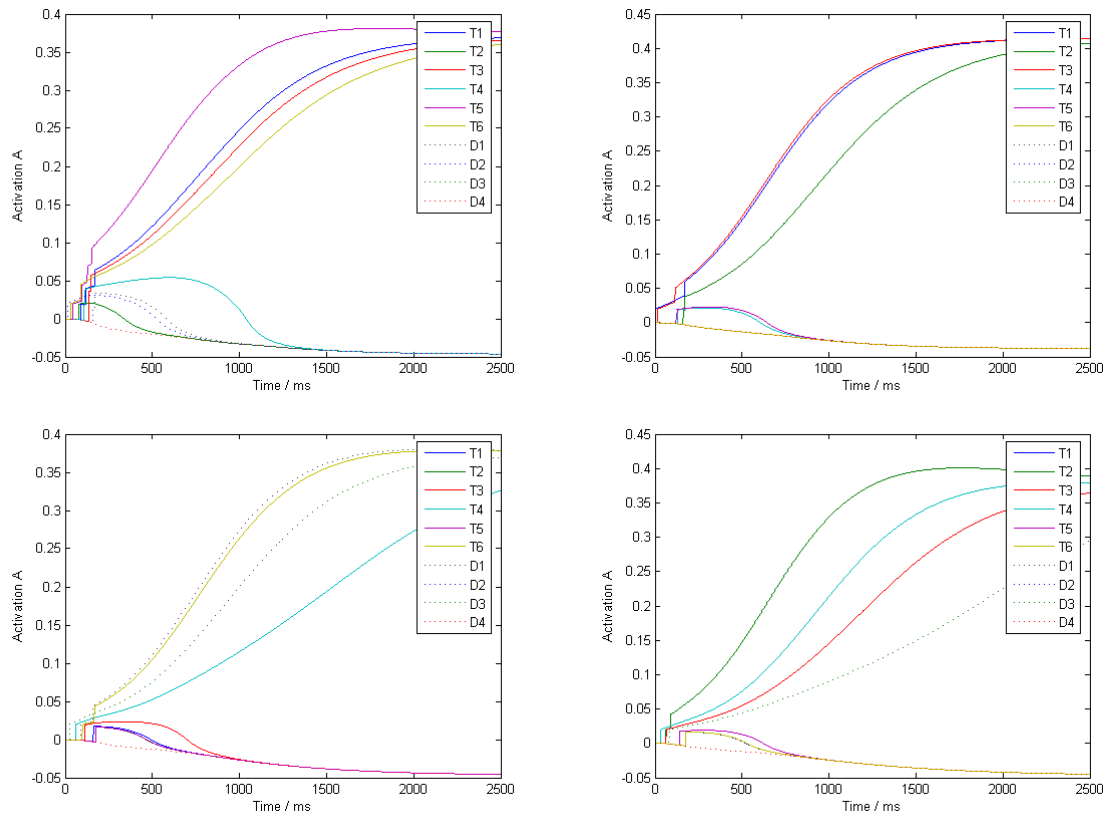


Figure 27: Dynamics of the Non-Unit Spike Model. Parameters are the ones mentioned in Table 2 for the NUSM. The four plots represent activation levels as a function of time. Experimental conditions for all four plots were: $T = 6$, $D = 4$ and $t_e = 200$ ms. The four simulations plotted were selected at random.

Compared to Figure 25 in Figure 27 we shall note that spikes now have a non-unit height ($\gamma^* = 2$), and further we note in an electrical engineering jargon that the ‘steady-state’ value of the activation level is only about 0.4 compared to the steady-state level in Figure 25 that was closer to 2.7. This change reflects the fact that now we have a different ratio between α^* and β^* .

10 The conservatory non-unit spike model (CNUSM)

10.1 Modifying the non-unit spike VSTM model

Starting from (9.1) we now have scaling factors for self-excitation, lateral inhibition, as well as for the input spikes. Scaling factors being α^* , β^* and γ^* . As the model appeared in (8.1) as well as in (9.1) lateral inhibition was simply sent from any active assembly to any other assembly (excluding the assembly itself).

Our intention with this chapter is to investigate what happens, if active assemblies in some way receive less inhibition, than do inactive assemblies; inhibition stemming from other active assemblies. The reason that we name the model the conservatory non-unit spike model, reflects the idea that the model in some sense ‘conserves’ activated inputs.

To model this phenomenon we introduce the function H , and hence expanding the model described in the previous chapter we obtain a modified model, described by

$$\frac{dA_x}{dt} = -A_x + \alpha^* F(A_x) - \beta^* H(A_x) \sum_{z \neq x} F(A_z) + \gamma^* G(v_x) \quad (10.1)$$

The newly introduced function H we define as:

$$\begin{aligned} H(A_x) &= h, & \text{for } A_x > 0 \\ H(A_x) &= 1, & \text{for } A_x \leq 0 \end{aligned} \quad (10.2)$$

For simplicity we shall only explore the case where $h = 0$. We note in this limit that H reduces simply to a ‘logically inverted’ version of the Heaviside step function.

The parameter h actually implements a sort of gating mechanism that is taken to the extreme by setting $h = 0$. The gate makes it harder for non-activated inputs to enter the VSTM. In contrast, activated inputs receive less or even no lateral inhibition, which means that once an input has been activated, it will either become very hard or in the extreme case even impossible to cancel its activation. In relation to biological neurons it is commonly known that neurons can sometimes act as so-called veto cells. These are neurons that, if they are activated, are able to veto other cells from firing, veto cells are single cells that can act as modulators of input. If the gating mechanism that we suggest does in fact exist in the human brain (in relation to the VSTM), it even would be possible to implement it on a single-neuron level! What we suggest is that an activated assembly, in the very moment when it is activated, cuts off inhibition stemming from any other activated assemblies. This could easily be implemented with a veto function, and possibly the mechanism acts on a mean field scale.

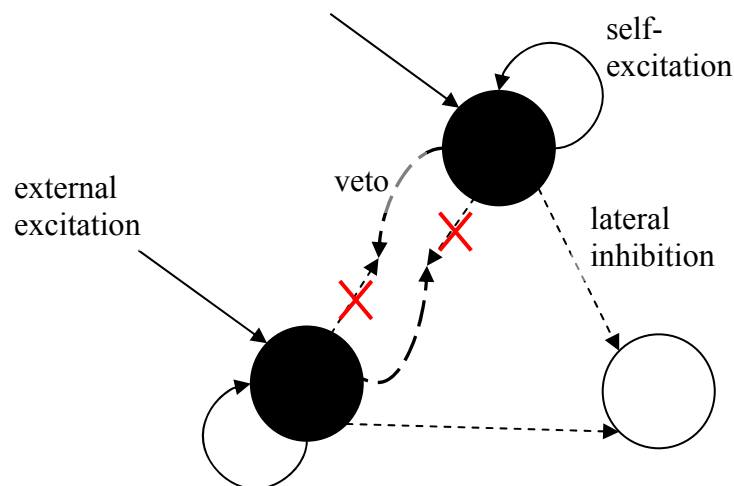


Figure 28: The act of the veto mechanism that in effect is implemented by the inclusion of the h parameter. We see that two neuron assemblies are activated via external excitation; these are shown as dark filled circles in the figure. The activated assemblies re-excite themselves via self-excitation, and further they try to inhibit all other assemblies from becoming active. As seen only non-activated assemblies (one of these is illustrated by a white circle) receive strong lateral inhibition from the activated assemblies. Assemblies that are already activated protect themselves with the built-in veto mechanism, and hence they only receive little or even no lateral inhibition from other activated assemblies. The veto mechanism is further described on p.85 in [9].

In relation to Figure 28 the parameter h can in fact be considered as the relative strength of the veto mechanism. When $h = 0$ this corresponds to the situation, where all lateral inhibition is ‘vetoed’ away. The Simulink implementation of the CNUSM (which also covers the USM, and the NUSM as special cases) can be found in Appendix B.

10.2 Shibuya & Bundesen’s data – Estimating parameters for the model

An advantage of the CNUSM model over the USM and NUSM is that in our case it is computationally more efficient to simulate. When $h=0$ we can use a simulation stop time as short as 200 ms. This corresponds to the point in time, where we do not have elements being encoded into VSTM, i.e. when all encoding rates are zero. The reason that 200 ms is enough to select the winners consolidated in VSTM is that once an element is activated, it will always stay activated due to the very strong veto protection mechanism ($h=0$).

As with the USM and the NUSM we also plotted a number of error maps for the CNUSM. In Figure 29 each map corresponds to a fixed value of γ^* ranging from 0.25 to 8. The plots in Figure 29 below were all produced in a similar way as described in section 8.4 above. Thus the error function we used was once again the one presented in equation (8.4). In Figure 29 we can note that it appears like if the error optimal regions are shaped as sort of an angular pond region. As γ^* is increased we note that β^* does generally also need to be increased, if one should stay in the error optimal region of the model’s parameter space.

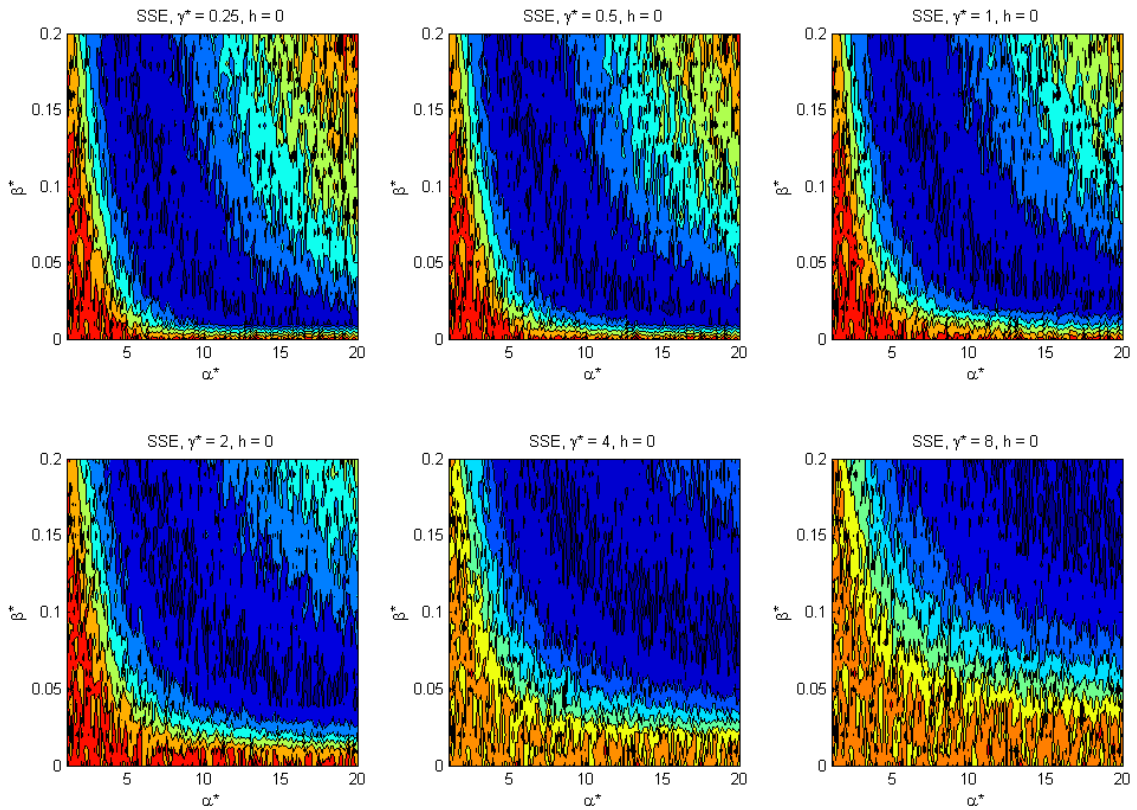


Figure 29: Cost function dependency as a function of α^* and β^* . Each window shows the cost function for a fixed value of $\gamma^* \in [0.25; 8]$. The dark blue areas represent the optimal parameter regions from an error point-of-view. A total of 200 trials were conducted for each setting of the model parameters.

In Figure 30 below we see six more error maps for the CNUSM. This time each map corresponds to a fixed value of α^* ranging from 1 to 6. Plotting procedures were similar to the procedures used with Figure 29. It appears from Figure 30 that the optimal ratio between the two parameters β^* and γ^* is largely independent of the setting of α^* , at least in the operated parameter domain. This latter finding actually comes as no big surprise to us. This we can say, since with the CNUSM once an input has first been activated it will always end up in VSTM. This statement is independent of the setting the activation (we assume $\alpha^* > 1$).

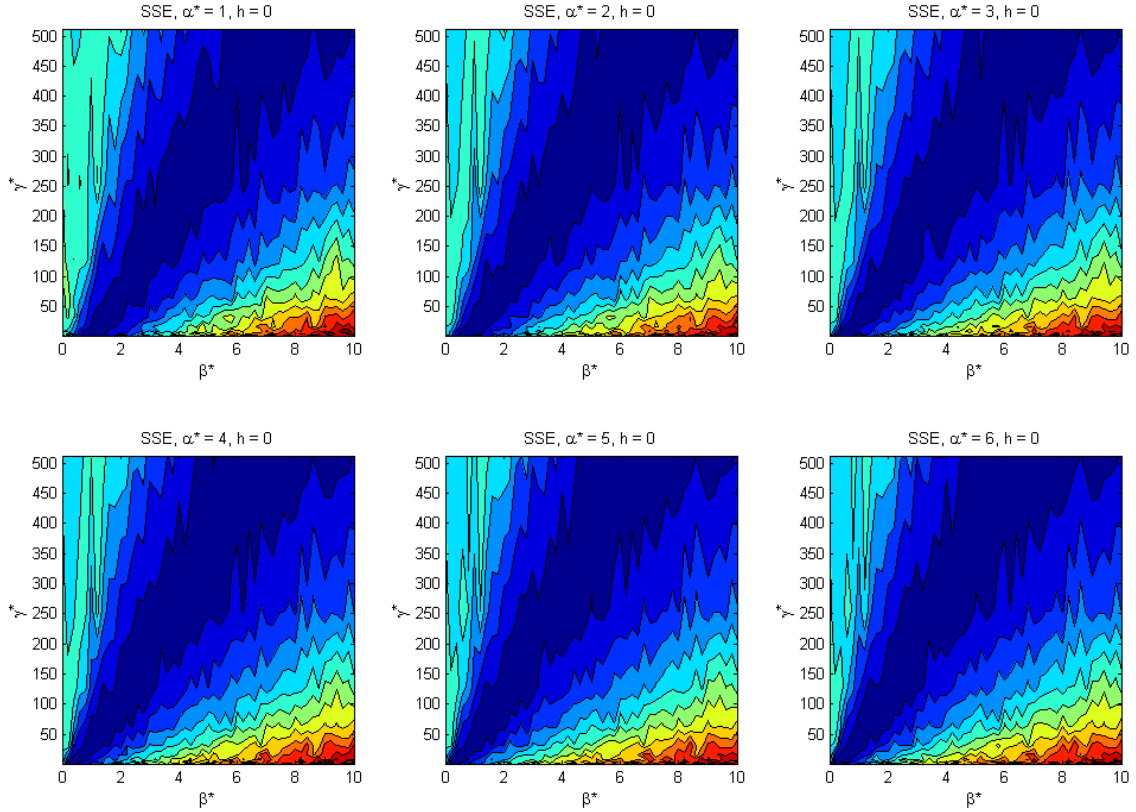


Figure 30: Cost function dependency as a function of β^* and γ^* . Each window shows the cost function for a fixed value of $\alpha^* \in [1;6]$. The dark blue areas represent the optimal parameter regions from an error point-of-view. A total of 200 trials were conducted for each setting of the model parameters.

10.3 Attentional dwell time modelling

With the NUSM one of the reasons for introducing γ^* was also to investigate, if this would enable the model to account for attentional dwell time phenomena. With the NUSM we were actually able to develop a dynamical behaviour, so that the second stimulus presented would be hindered from entering VSTM. However the parameter set (α^* , β^* and γ^*) that was needed to obtain this kind of behaviour was certainly not comparable to the parameter values that we had used in fitting the WR & PR data in [38]. So, in the end our problem was that we were not able to determine parameters that would give us both a strong attentional blink effect and a storage capacity as high as in the order of four visual elements. An important reason for implementing the CNUSM was that we thought it might avoid the latter mentioned problem. In fact we found that CNUSM was significantly better at modelling attentional dwell time experiments than the NUSM.

Having already obtained a set of parameters for the CNUSM from fitting the model to the WR & PR experiment of Shibuya & Bundesen's, we further hand-tuned the model to fit also the attentional dwell-time experiment plotted in Figure 11. In fact we ended up in a rather different parameter region than we had initially assumed (cf. Figure 29). The results we obtained were remarkably good, both in the attentional dwell time situation – but actually also in the WR & PR situation!

In Figure 31 below the CNUSM was applied to simulate the experimental conditions described in section 3.3. The empirical results from Ward, Duncan & Shapiro's studies were shown in Figure 11. Note that close resemblance we have between Figure 31 and Figure 11. The parameters are the ones that we found via hand-tuning, and these are also the ones listed in Table 2.

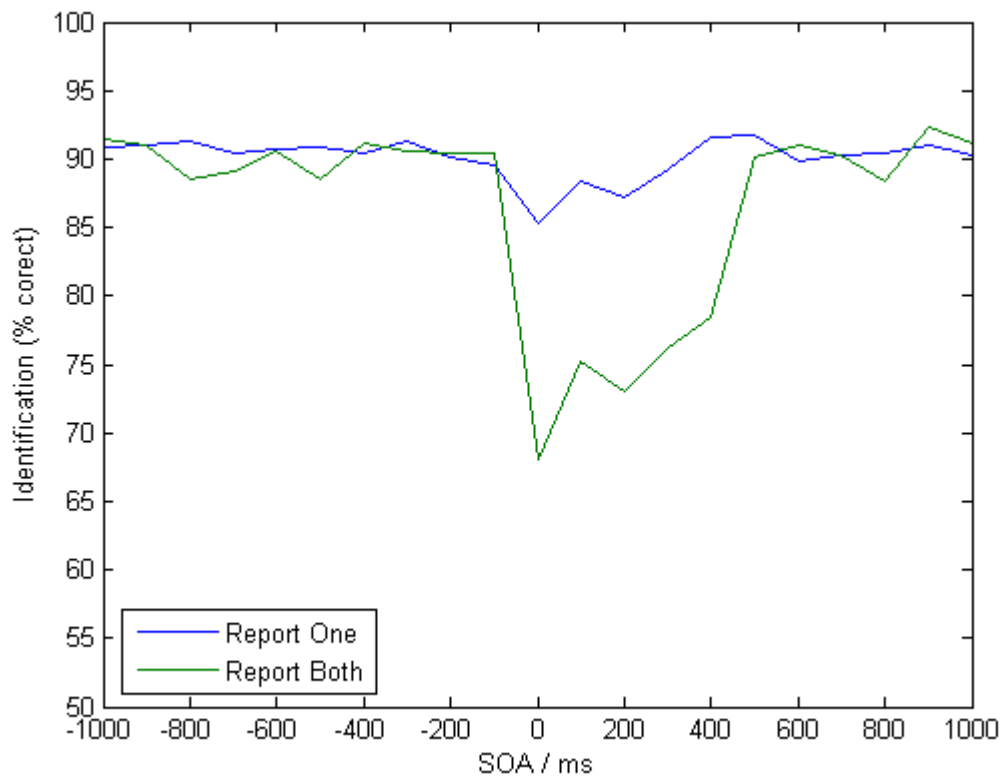


Figure 31: Simulation corresponding to attentional dwell time experiment by Ward, Duncan & Shapiro (1996). The figure should be compared with Figure 11. Model parameter settings are the ones listed in Table 2 for the conservatory non-unit spike model. A similar number of trials per experimental condition as in the empirical experiment were used in the simulation.

With Figure 31 we note that there are only small miss-matches between the rates of identification for the model and the empirical identification rates shown in Figure 11. The most noticeable mismatch occurs at lag-0, i.e. when both of the characters are presented simultaneously. Here the model predicts a rather low identification rate, compared to the identification rate at later lags. A general concern when fitting attentional dwell-time data with our model is that we must ask ourselves, if we do really consider all effects caused by the mask, see section 2.7.

Interestingly we see that our model predicts sparring of the input at lag-1. This also appears to be an important finding in relation to attentional dwell-time studies.

The model parameters applied to produce Figure 31 were also applied to produce the results shown in Figure 32.

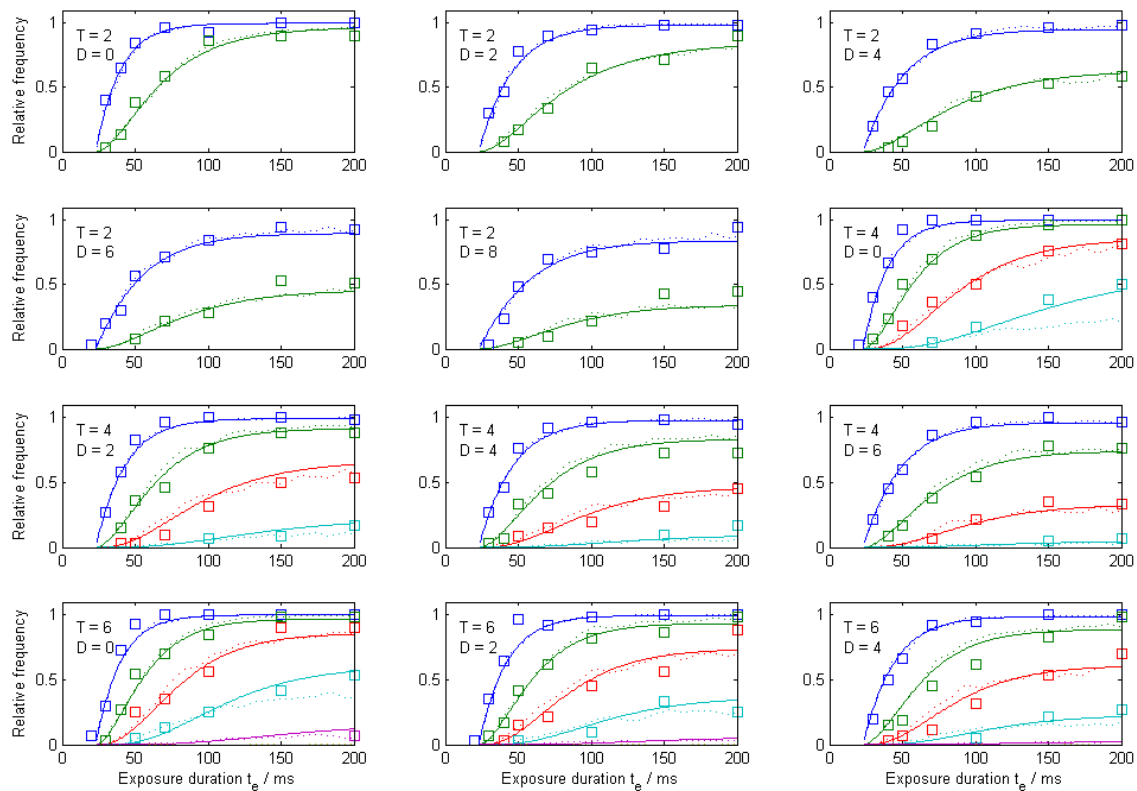


Figure 32: Simulation corresponding to WR & PR experiment by Shibuya & Bundesen (1988). The empirical data are shown as squares. Dotted lines are from the CNUSM fitting and solid lines are the fit by the MCMC, which is included simply for comparison. Again we applied the CNUSM parameter settings of Table 2.

As is seen from Figure 32 the CNUSM actually fits the WR & PR data of Shibuya & Bundesen remarkably well. Since we used $\gamma^* = 150$, we do not know much about the region of the parameter space, in which the model is located, see Figure 29. For this reason we once again plotted an error map similar to the ones that we have already seen. The only difference is that now $\gamma^* = 150$, and further we expanded the (α^*, β^*) -plane.

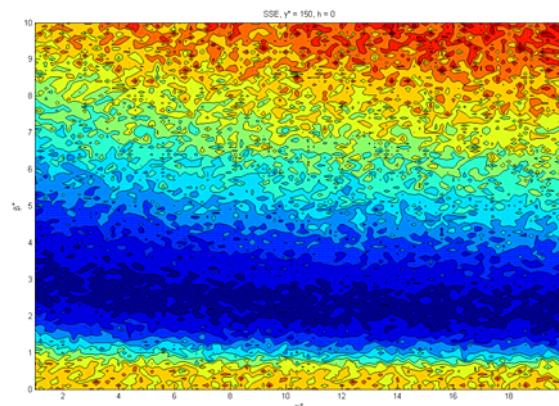


Figure 33: Error map for the CNUSM model. The parameters we used in order to produce Figure 31 and Figure 32 were $\alpha^* = 1.2$, $\beta^* = 2.6$, $\gamma^* = 150$ and $h = 0$. Notice that the α -axis begins with $\alpha^* = 1$ which is the smallest value of α^* , where we have sustained solutions for the activation levels, see [37], section 2.4. A total of 200 trials were conducted for each setting of the model parameters.

The new error map for $\gamma^* = 150$ is shown in Figure 33. As we can see the parameter set we applied to produce Figure 31 and Figure 32 seems to be in a close-to-optimal region of the map. The error optimal region almost appears as a thick (dark blue) horizontal line, where the value of β^* is only slightly dependent upon the value of α^* . This again confirms our notion in section 10.2 that with the CNUSM and with $h = 0$ the setting of α^* is actually not so important in order to model WR & PR data.

10.4 Dynamics of the model

Concerning the dynamics of the model the notions in section 8.5 for the USM applied for the NUSM as well. Many of these notions also apply for the CNUSM.

In Figure 34 below we see once again activation levels depicted as a function of time for four identical simulations. Again experimental conditions were similar to the ones used in section 8.5 for USM.

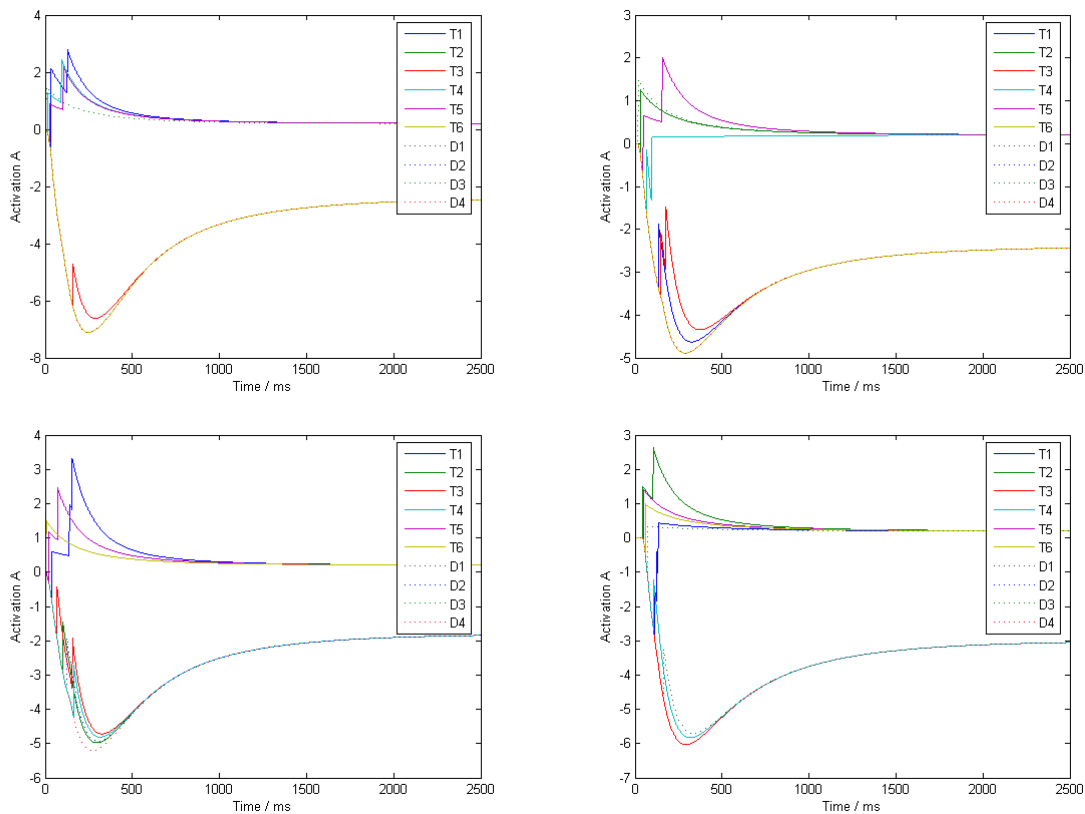


Figure 34: Dynamics of the Conservatory Non-Unit Spike Model. Parameters are the ones mentioned in Table 2 for the CNUSM. The four plots represent activation levels as a function of time. Experimental conditions for all four plots were: $T = 6$, $D = 4$ and $t_e = 200$ ms. The four simulations plotted were selected at random.

In Figure 34 above we note that spikes are very importantly assigned a considerable height ($\gamma^* = 150$) compared to the ‘steady state value’. The steady state value is in fact

very close to, but above zero. Also we see that β^* is rather large compared to α^* , since non-activated visual elements (elements with a negative activation level) are strongly inhibited by the already activated elements. Even three or four visual elements in 'steady-state mode' (i.e. in the last part of the plot where the activation levels are converging) seem to be able to guarantee a strong level of inhibition in the network. The inhibition makes it more difficult for later presented input to enter VSTM. Note that in the four plots in Figure 34 the activation level of the suppressed input usually takes a dip with a negative peak after about 250 ms. The dip seems to last for some 1000 ms, which is in a very fascinating way very close to the duration of the attentional dwell time, see Figure 11. Note that at the peak of the dip it becomes very much harder for suppressed input to 'reach the surface' and become encoded into VSTM.

In fact with the CNUSM we suggest that the attentional blink arises due to a sophisticated overshoot effect. This term is directly taken from electrical engineering control theory, where an overshoot is defined as: *the ratio between the largest value of the response and the stationary ending value (i.e. the steady state value)*.

PART III:
RESULTS – FIT TO DATA

11 Results

11.1 Fitting the storage capacity of VSTM in a whole report study

Below we show the fitting results obtained with the binomial and the hypergeometric distribution (see chapter 4). Note that the empirical distribution is followed most closely by the hypergeometric distribution. The binomial distribution produces a less close fit to the empirical distribution, and generally it is too broadly distributed. In general the best fits are obtained with $K = 5$, which also appears from the *SSE* in Table 1.

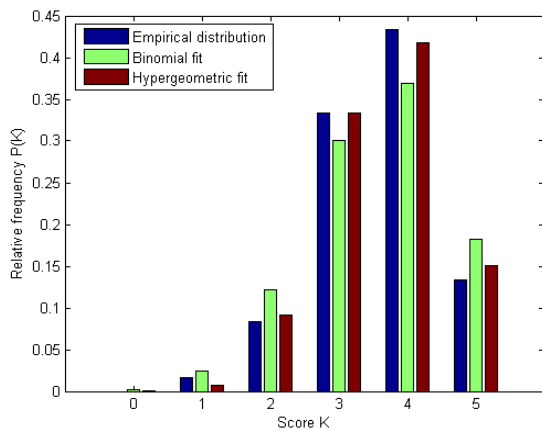


Figure 35: Fitting with $K = 5$

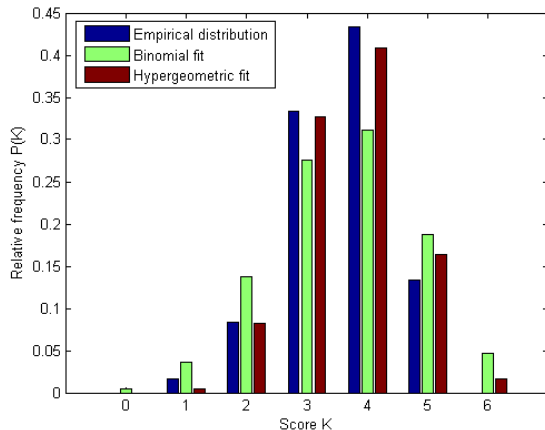


Figure 36: Fitting with $K = 6$

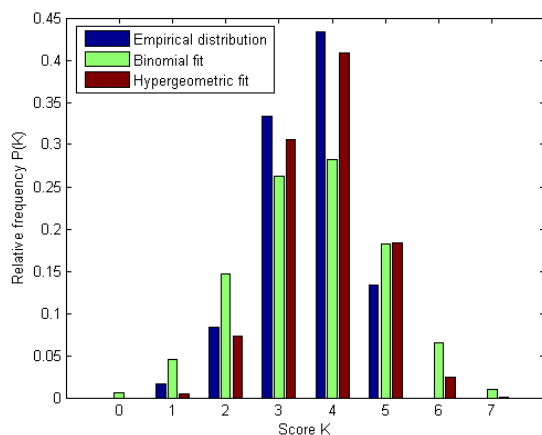


Figure 37: Fitting with $K = 7$

K	Binomial	Hypergeometric
5	n 5	n_{sa} 13
	p_s 0.71	K_{tot} 18
	SSE $9 \cdot 10^{-3}$	SSE $7 \cdot 10^{-4}$
6	n 6	n_{sa} 8
	p_s 0.60	K_{tot} 13
	SSE $3 \cdot 10^{-2}$	SSE $2 \cdot 10^{-3}$
7	n 7	n_{sa} 7
	p_s 0.52	K_{tot} 13
	SSE $4 \cdot 10^{-2}$	SSE $5 \cdot 10^{-3}$

Table 1: Fitting results from the three fits.

11.2 Preliminary testing of Usher & Cohen's original model

Figure 38 shows the development in mean memorial capacity as filled contour level plots (rounded to the nearest integer). Figure 39 shows linear fits to a contour level curve fixed at a mean memorial capacity of 3.74. (This was the K -value reported for subject MP in [38]. For the two plots shown in the bottom of the latter figure we note that neither intersection nor slope really makes any sense when $N_{na} < 4$. This is rather logical, since we should at least have enough inputs presented so that we can actually fill the memory, otherwise we can not really test the memorial capacity.

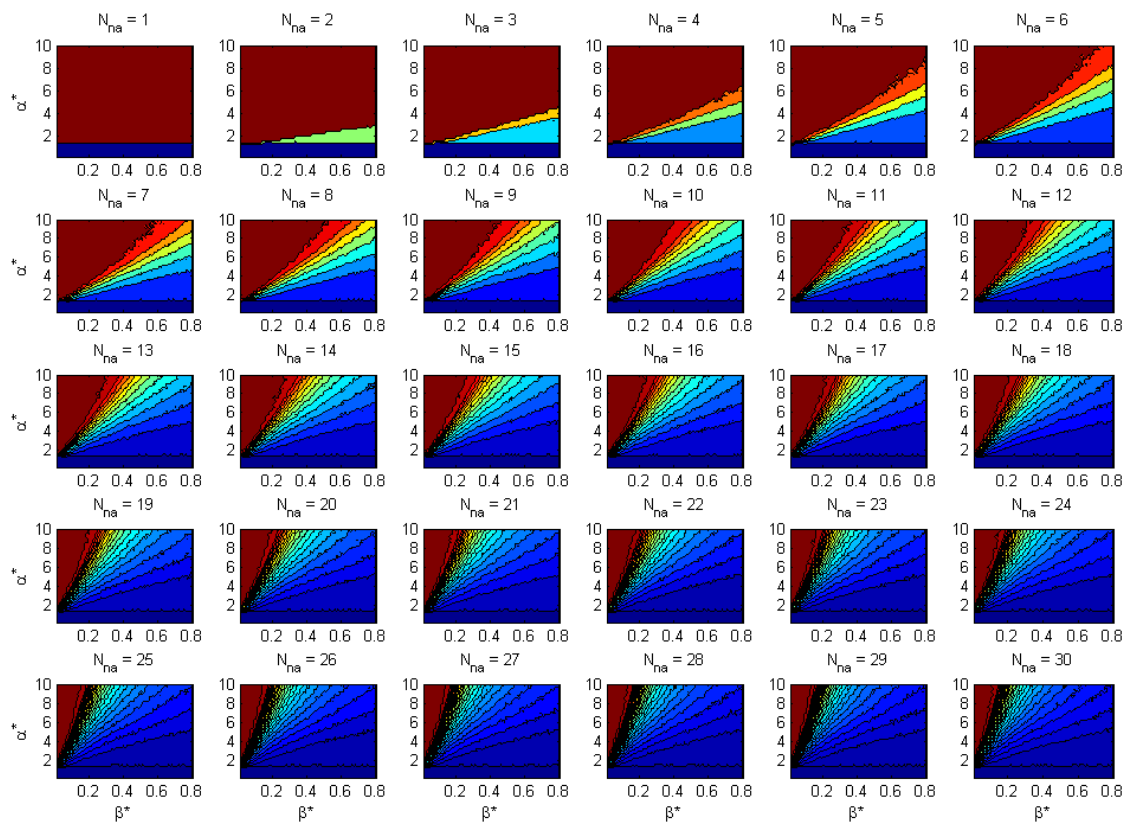


Figure 38: Filled contour plots displaying mean memorial capacity as a function of different α^* - β^* -settings. Deep red demarks a region where the mean memorial capacity is N_{na} . Independent of the setting of N_{na} we see that each window has a region for α^* -values smaller than one that is deep blue. The deep blue regions represent parameter spaces, where the mean memorial capacity is zero.

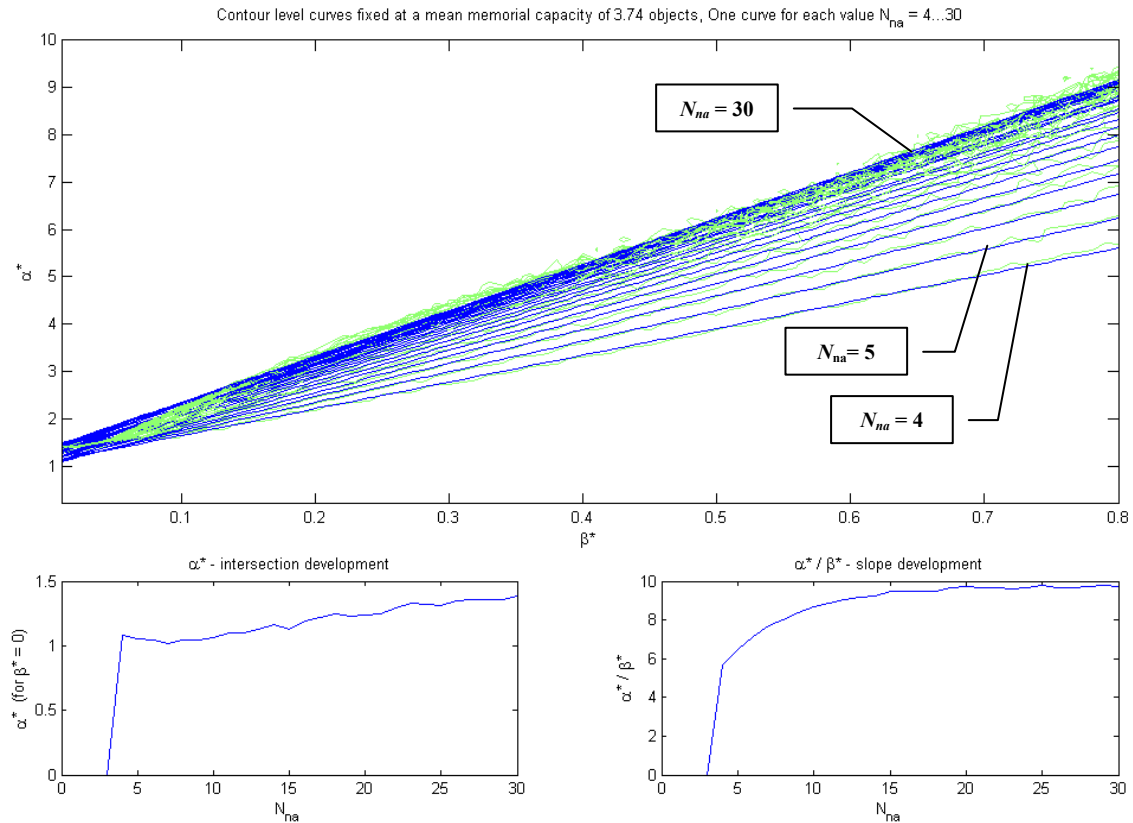


Figure 39: Upper window shows development of α^* - β^* curves for an increasing number of neuron assemblies N_{na} (shown in solid green lines). Linear fits to the α^* - β^* curves are shown as solid blue lines. Bottom left window shows α^* -intersection for $\beta^* = 0$, and bottom right window shows slope development.

From Figure 38 and Figure 39 we see that we were correct in our hypothesis from section 5.3. Especially from the latter figure it clearly appears that in Usher & Cohen's original model the model parameters α^* and β^* can not be optimized independently of the number of neuron assemblies N_{na} . The fact that we changed Usher & Cohen's model in such a way that we do not have this problem with the spike model, is reason enough for us not to consider this weakness of Usher & Cohen's model in any further detail.

11.3 The fit to Shibuya & Bundesen's partial and whole report data

Whether a model capable of producing better fits is actually a better model can only be determined from calculating a measure for the feasibility of the model. Typically such a measure will include a term that considers the number of free parameters N_{fp} in the model, this is the case with both Akaike and Bayes Information criteria, *AIC* and *BIC*.

Below the results from calculating the *AIC* and *BIC* are shown, for this purpose we first calculated the negative log-likelihood *NLL*. The results show that including all components K from 1 to 5 does yield a better model. For the three mixture models the optimum parameters that we have indicated are the parameters that we found using

nonlinear constrained optimization. For the rest of the models, i.e. for the three spike models the parameters indicated are more or less found by trial and error. The parameters that were tested for these models were mainly picked based on our knowledge of the simplified error function indicated in Figure 24, Figure 26 and Figure 29. With the CNUSM we managed to find a more optimal set of parameters outside the parameter region plotted in Figure 29, this also appears from Table 2. We note that there is generally a fine agreement between the two information criteria. This is further illustrated below in Figure 40. All measures for the three spike models were calculated based on a simulation setup with 500 trials per experimental condition.

PARAMETER	Mixture Models			Spike Models		
	2CMM	2CMM ₂	MCMM	USM	NUSM	CNUSM
C	48.7 s⁻¹	55.7 s⁻¹	61.5 s⁻¹	(61.5 s⁻¹)	(61.5 s⁻¹)	(61.5 s⁻¹)
t_0	19 ms	23 ms	23 ms	(23 ms)	(23 ms)	(23 ms)
α	0.40	0.41	0.367	(0.367)	(0.367)	(0.367)
$P_{K=1}$			0.04			
$P_{K=2}$			0.12			
$P_{K=3}$	0.26	0.07	0.26			
$P_{K=4}$	0.74	0.93	0.53			
$P_{K=5}$			0.05			
α^*				4	5	1.2
β^*				0.09	0.1	3.6
γ^*					2	150
h						(0)
MEASURE						
N_{fp}	4	4	7	5	6	6
SSE	1.3950	1.8326	1.1015	1.6307	1.6462	1.6013
NLL	5220	5192	5108	5189	5204	5176
AIC	10448	10392	10230	10388	10420	10364
BIC	10475	10419	10277	10431	10461	10405

Table 2: Results from fitting the various models to Shibuya & Bundesen's data set (subject MP). 'Free parameters' are written in bold and 'fixed' parameters are written in parentheses. Notice that the spike models have inherited the three upper-most parameters from the MCMM. Further note that for the conservatory non-unit spike model h was fixed to zero for simplicity.

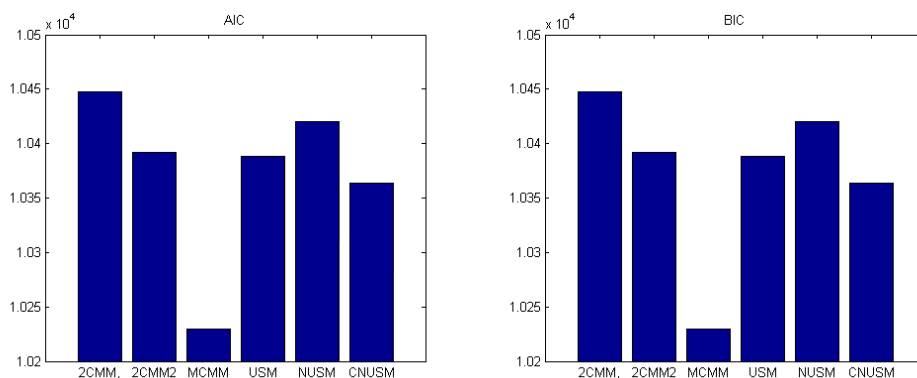


Figure 40: Model error charts - left: AIC , right: BIC . We see that the two measures agree; with the multi-component mixture model being much more optimal than Bundesen's 2-component mixture model. According to both of the criteria our spike models all perform better than Bundesen's model; with the best of the spike models being the conservatory non-unit spike model.

The two histograms in Figure 40 above appear to be almost identical. This we can note is due to the fact that the data set of Bundesen is actually a very large dataset, where the number of trials conducted was in fact very large compared to the number of free parameters that we have in the various models.

11.4 The fit to attentional dwell time data

We have not calculated any model errors, when it comes to fitting attentional dwell time data. However in Figure 31 we saw that the conservatory non-unit spike model (CNUSM) was in fact capable of fitting the empirical data of Ward, Duncan & Shapiro shown in Figure 11. This, worth noting, with the same set of parameters that we also used to fit the complete WR & PR experiment of Shibuya & Bundesen! The simple reason that we decided not to make any quantitative assessment of the model (in respect to fitting attentional dwell time data) is that we have not found any equivalent theoretical model that we could use for comparison. Hence to our knowledge, our model is deemed to be the best on the market!

**PART IV:
DISCUSSION, CONCLUSION
AND FUTURE WORK**

12 Discussion

12.1 Opening up the black box of visual attention

A good strategy for solving a jigsaw is first to get hold of the corners and then try to build up the boundaries, before the rest of the puzzle is to be solved. Perhaps a similar strategy is appropriate, when it comes to understanding how visual attention works. In a puzzle corners can also be understood as unique pieces, i.e. corners are extreme cases since they only connect to two rather than three or four other pieces.

We think that WR & PR experiments can be seen as extreme cases of visual report studies, in the sense that all elements are presented simultaneously and for an equal amount of time.

We speculate that a feasible way of reaching a better understanding of visual attention and VSTM is to start out with a good model for WR & PR and then try to generalize the model to cover other pieces of evidence like results from attentional dwell time experiments. It seems a fair assumption that an overall model for visual attention must be able to explain results from as many different types of visual experiments as possible. Since WR & PR and attentional dwell time studies are both considered important types of experiments probing visual attention, it seems logical to search for a common model.

When troubleshooting a good engineer should always try to test as small a part of the system as possible at a time. In connection with cognitive models it seems particular difficult to follow this typical engineering approach of isolating the problem, simply because cognitive processes are neurobiologically not well-defined. Further the various regions of the brain are often heavily interconnected and therefore potentially could take part in the cognitive processes that we are investigating.

12.2 Experimental design: Fokker-Planck versus Monte-Carlo

A stochastic differential equation (SDE) is defined as a differential equation in which one or more of the terms are stochastic processes thus resulting in a solution, which is itself a stochastic process [42].

Equations (8.1), (9.1) and (10.1) all belong to the general set of coupled first-order SDEs implying that theoretical methods for solution do actually exist.

Though other methods for reaching a solution do exist, the main method of solution would be to use the Fokker-Planck equation, which provides a deterministic equation satisfied by the time dependent probability density.

Using Fokker-Planck (FP) our aim would be to reach a deterministic equation that satisfies the probability density $p(A_1, A_2, A_3, \dots, A_N)$ of the N activation variables: $A_1, A_2, A_3, \dots, A_N$.

Unfortunately the computational burden of solving the FP equations is highly dependent upon the dimensionality of the probability space that we are trying to explore. In fact the computational complexity rises with the power of the dimensionality, i.e. with the power of N .

In practice 1D and 2D FP equations are the computationally most feasible to solve. Progressing to 3D FP equations these can be solved with COMSOL's FEMLAB toolbox. Going to even higher dimensionalities of FP equations, the random access memory requirements of the computations often grow to considerable heights. A 4D solution often could require in the order of GB's of RAM. Further, to our knowledge, no standard simulation tools exist for solutions of problems with order of 4 dimensions or more, thus one would need to derive and program the solution from scratch.

Referring to both Figure 4 and also Cowan, [10] the 'magic number' of visual objects is about 4, so it seems fair that at least $N=5$ neuron assemblies would be appropriate in order to reach a descriptive solution of all potential outcomes of VSTM capacity K .

Because of the many complications with solving the Fokker-Planck equations we chose an alternative tool for exploring the equations, namely Monte-Carlo-like simulations. The marked advantage of Monte-Carlo simulations is simply that computationally they are much less time-consuming.

12.3 Binomial and hypergeometric fits to whole report data

The binomial distribution fits the empirical distribution quite nicely, however it seems that the true distribution is closer to bimodal (with peaks at scores 3 and 4) than is the binomial distribution. The hypergeometric distribution generally contains more 3's and 4's, and less 2's and 5's than does the binomial distribution, and hence overall it is more successful at fitting the data. As seen from the summed squared errors, *SSEs*, the fit of the hypergeometric distribution actually outperforms the binomial distribution regardless of which maximal score ($K = 5, 6, 7$) that we choose. Though we generally obtain the best fit when $K = 5$, it is an interesting fact that setting the maximal score to 7 leads us to an optimal solution, where the number of storing attempts n_{sa} is actually down to exactly 7! In this situation we could imagine that we have one storage attempt for each target being presented (7 targets were presented). What we call the total number of storage scenarios K_{tot} seems harder to translate or interpret conceptually; however it does anyway appear that the number must in some way relate to a phenomenon that tends to inhibit all targets from becoming encoded into VSTM (i.e. making it more difficult to reach a memory of K visual elements).

12.4 Mixture model fits to whole and partial report data

Shibuya & Bundesen obtained very impressive fits to the empirical data in [38] using a simple two-component mixture of the FIRM. Generally the subject dependent parameters in their model can be determined in a computationally efficient way.

However based on a ML consideration and without the two computational assumptions reported in section 6.3, Shibuya & Bundesen would simply have had to reject their model.

Actually it showed out that Shibuya & Bundesen's assumptions had been chosen very conveniently in order for them to justify the use of a mixture model. When we applied the single more consistent assumption mentioned in section 6.3, we saw that we obtained a set of fitted parameters that was quite different from the set obtained by Shibuya & Bundesen. With the 2CMM₂ we found the mixing coefficient $P_{K=4} = 0.93$, which suggests that simply setting the storage capacity $K=4$ (instead of using $K=3,4$) would not make that big of a difference. Further with the 2CMM₂ we found $t_0 = 23$ ms, which was actually the same value that we obtained, when we fitted the MCMM.

The multi-component mixture model was proposed partly in order to avoid the assumption mentioned in section 6.3 and partly in order to model the broader distribution of the score j that empirically have been observed, see Figure 4. The MCMM generally does a better job at fitting the data, and according to measures like *AIC* and *BIC*, which also considers the numbers of model parameters, we can indeed conclude that it is a better model than both the 2CMM and the 2CMM₂.

12.5 Spike model fits to WR & PR and to attentional dwell time data

The unit spike model is a subclass of the non-unit spike model, which is again a subclass of the conservatory non-unit spike models.

The model parameters of the spike models are computationally rather demanding to determine compared to the parameters in the above mentioned mixture models. This is due to the fact that the spike models are stochastic models with a dynamic nature, hence rather than calculating everything at once we need many computational steps to check how the simulated system evolves.

In section 8.2 we provided visual as well as intuitive evidence that the model parameters C , t_0 and α can in a feasible fashion be adopted from a fit with the MCMM. If the parameters are directly adopted, this certainly limits the number of parameters that have to be optimized when fitting the spike models! Alternatively, if computational resources were less sparse, the parameters known from the MCMM could be used as appropriate initial values when optimizing the model parameters of the spike model.

Usher & Cohen's model is a model that is able to select, which winners should be encoded into memory. The stochastic nature of this winners-take-all kind of nature is

ensured by the inclusion of an additive noise term. We quickly discovered that such a stochastic nature could also be inducted by the use of Poisson distributed input spikes instead of additive noise. Hence, implementing our model with Bundesen's TVA we went for Poisson distributed input spikes, and for simplicity we did not include an additive noise term, when we proposed the USM.

The USM proved successful at fitting the WR & PR data in [38]. Compared to Bundesen's own 2CMM which treated extreme scores (scores > 4) as outliers, our USM encompasses these as natural consequences of the internal dynamics.

The NUSM was introduced to investigate the effect of varying the amplitude of the input spikes. By trial and error we found that a model where the spike amplitude is 2 yields a slightly closer fit to the data in [38] compared to the fit that we obtained with the USM (spike amplitude of 1). Thus it seems that the influence from varying γ^* can not be completely compensated by changing the values of α^* and β^* . We shall note that it is primarily for large exposure durations (from 100 to 200 ms) that the setting of the parameters α^* , β^* and γ^* has a great deal to say, since for shorter exposure durations it seems to be the underlying nature of the Poisson processes that control the game. This is due to the fact that at this early state in the process, the inhibition in the network is so small that it does actually only has little influence on the setting of the storage capacity. One of the reasons for introducing γ^* was also to investigate, if this would enable the model to account for attentional dwell time phenomena. We were actually able to develop a dynamical behaviour, so that the second stimulus presented would be hindered from entering VSTM. However the parameter set (α^* , β^* and γ^*) that was needed to obtain this kind of behaviour, was certainly not comparable to the parameter values that we had used in fitting the WR & PR data in [38]. So in the end our problem was that we were not able to determine parameters that would give us both a strong attentional blink effect and a storage capacity as high as in the order of four visual elements.

The CNUSM was developed to address the latter mentioned problem. The parameter h was introduced so that now we could setup different levels of inhibition for activated and non-activated assemblies. For simplicity we only investigated the case when h was fixed to 1, i.e. the extreme situation where only non-activated assemblies receive inhibition. With the CNUSM we were able to determine values α^* , β^* and γ^* that would give us a model capable of demonstrating the attentional blink effect (as seen in Figure 11) and at the same time being able to fit the WR & PR data in [38].

Here we shall note that the WR & PR data were for a single subject study, while the attentional dwell-time data were actually averaged over many subjects as previously described. The optimal situation probably would have been, if all data had been collected from experiments with the same single subject, however to our knowledge such data sets do not exist at the time being (we hope that this thesis will act as a motivation for conducting such type of combined experiments with single subjects).

As we see in Figure 34 the peak of the inhibitory dip taken by non-activated elements is directly related to how many elements managed to fire a spike and go into the region of positive activation. In the four plots shown the peak of the dip varies between -5 and -7. This means actually that the number of visual elements that have already entered VSTM

is very important for the level of lateral inhibition in the network. In Figure 12 (in the 'Report All' condition) we have a situation where two elements have already been stored in VSTM when the third element is presented. Empirically we see that the fact that two rather than one elements are already stored in VSTM when the next element is displayed is directly related to an expanded attentional blink effect. We believe our model would be able to mimic exactly such kind of behaviour however the experiment remains for future investigations.

13 Conclusion

In this thesis we have reached a number of important conclusions in our continuing efforts to implement a neural network model capable of simulating the behaviour of the visual short-term memory experienced during psychological studies of visual attention.

- First we showed that whole report data can be fitted by a binomial distribution, as expected. However, we also discovered that even closer fits can in fact be obtained with the hypergeometric distribution.
- Then we improved Shibuya & Bundesen's two-component mixture model by expanding it to a multi-component mixture model with five rather than two components.
- Further we showed that the parameters estimated with this multi-component mixture model can in a feasible fashion be adopted by the spike model.
- Finally we developed the spike model by integrating Bundesen's theory of visual attention with the short-term memory model of Usher & Cohen. The spike model was developed in three consecutive stages, and in the final stage the model was able to give account of both a comprehensive set of whole and partial report data, as well as it was able to predict the so-called attentional blink that is perhaps the most important finding in attentional dwell time studies.

Addressing our experimental design we believe Monte-Carlo simulations is in fact one of the best tools for learning more about the stochastic differential equations that govern our dynamic model; equations we think model the behavior of the VSTM very well.

In conclusion we note that our new dynamic spike model of visual attention and visual short-term memory is able to account for the complete set of data from whole and partial report experiments. Where the previous account by Shibuya & Bundesen [38] treated extreme scores as outliers, the new model encompasses these as natural consequences of the internal dynamics.

Further we note that the spike model explains VSTM capacity and consolidation as the result of a dynamic process rather than as a static store, and importantly also with our new model, the storage capacity K is still kept independent of both the processing capacity C , and independent of the attentional set of the subject (the β - and π -values in Bundesen's TVA).

14 Future work

In future studies we wish to explore the model's ability to explain the dynamic consolidation in VSTM found in temporally extended paradigms such as the attentional blink paradigm and studies of attentional dwell time (e.g. [40]). Here consolidation in VSTM is strongly dependent on competition between items already encoded into VSTM and visual items presented at a later point in time. This competitive process follows naturally from the dynamic architecture of the present model.

In future work it would be interesting to explore, what can be learnt from a Fokker-Planck formulation of our stochastic differential equation. Especially if one would be able to predict the distributional function that governs j , one could hope to gain enough insight to be able to relate the distributional parameters to parameters in the differential equation.

Further it would be interesting to expand the model to cover multiple categorizations of the same object. In this way the model should be expanded to cover studies, where subjects are instructed to report two features of the same visual element (e.g. letter identity as well as color). In fact the TVA block has already been implemented in such a way that it permits us to have multiple categorizations of the same object in future studies.

As we have seen that whole report data can be approximated very closely it would be interesting to know if parameters could be determined that would make our spike VSTM model converge towards the hypergeometric distribution. Evidence for such behavior possibly would demand a more analytic approach to the model.

We implemented the spike models and tested them against WR & PR data. Unfortunately we did not have time or computational resources to perform a comprehensive optimization of the CNUSM; this however would be an interesting topic for future work. Indeed if we compare the parameters mentioned in Table 2 for the CNUSM with Figure 33 it seems that chances are very good that more optimal parameters can be obtained, at least when it comes to fitting the WR & PR data.

We have reason to believe that the CNUSM is very likely able to mimic the 'expanded' attentional blink effect observed in Figure 12 during the 'Report All' condition; the verification of this assumption remains for future work.

Appendix A FIRM equations – whole and partial report

Parameters in FIRM

w_d : Attentional weight of a distractor.

w_t : Attentional weight of a target.

$\alpha = \frac{w_d}{w_t}$: Relative weight of distractors and targets.

K : Capacity of the VSTM [elements].

C : Total processing capacity [elements/s].

T : Number of targets in the display.

D : Number of distractors in the display.

Processing rate of a target:

$$\begin{aligned}v_t &= \frac{w_t}{w_t T + w_d D} C \\&= \frac{w_t}{w_t T + \alpha w_t D} C \\&= \frac{C}{T + \alpha D}\end{aligned}$$

Processing rate of a distractor:

$$\begin{aligned}v_d &= \frac{w_d}{w_t T + w_d D} C \\&= \frac{\alpha w_d}{w_t T + \alpha w_t D} C \\&= \frac{\alpha C}{T + \alpha D} \\&= \alpha v_t\end{aligned}$$

Probability distributions

Probability distribution function for target processing:

$$F(t) = 1 - \exp(-v_t t)$$

Probability density function for target processing:

$$f(t) = v_t \exp(-v_t t)$$

Probability distribution function for distractor processing:

$$G(t) = 1 - \exp(-v_d t)$$

Probability density function for distractor processing:

$$g(t) = v_d \exp(-v_d t)$$

Available processing time:

$$\tau = t_e - t_0$$

where t_e is the exposure duration and t_0 is the longest ineffective exposure duration.

Probabilities

Probability of a score of j (targets reported) from a display containing T targets and D distractors exposed for t_e seconds:

$$P(j; T, D, t_e) = P_1 + P_2 + P_3$$

P_1 is the probability that the score equals j and the total number of elements (targets and distractors) entering VSTM is less than K . The number of distractors entering VSTM is denoted by m and $m \leq \min(D, K-j-1)$. If $j=K$, $P_1=0$; otherwise:

$$\begin{aligned} P_1 &= \binom{T}{j} [F(\tau)]^j [1 - F(\tau)]^{T-j} \cdot \sum_{m=0}^{\min(D, K-j-1)} \binom{D}{m} [G(\tau)]^m [1 - G(\tau)]^{D-m} \\ &= \binom{T}{j} [1 - \exp(-v_t \tau)]^j [\exp(-v_t \tau)]^{T-j} \cdot \sum_{m=0}^{\min(D, K-j-1)} \binom{D}{m} [1 - \exp(-v_d \tau)]^m [\exp(-v_d \tau)]^{D-m} \end{aligned}$$

After inserting v_t and v_d given by α , C , T , and D :

$$\begin{aligned} P_1 &= \binom{T}{j} \left[1 - \exp\left(-\frac{C}{T + \alpha D} \tau\right) \right]^j \left[\exp\left(-\frac{C}{T + \alpha D} \tau\right) \right]^{T-j} \cdot \\ &\quad \sum_{m=0}^{\min(D, K-j-1)} \binom{D}{m} \left[1 - \exp\left(-\frac{\alpha C}{T + \alpha D} \tau\right) \right]^m \left[\exp\left(-\frac{\alpha C}{T + \alpha D} \tau\right) \right]^{D-m} \end{aligned}$$

P_2 is the probability that the score equals j and the total number of elements equals K and the K -th element entering the VSTM is a target. The number of distractors entering VSTM denoted by m is always $K-j$. If $j=0$, or $j < K-D$, $P_2=0$; otherwise:

$$\begin{aligned}
P_2 &= \int_0^{\tau} \binom{T-1}{j-1} [F(t)]^{j-1} [1-F(t)]^{T-j} \cdot \binom{D}{m} [G(t)]^m [1-G(t)]^{D-m} \cdot \binom{T}{1} f(t) dt \\
&= \binom{T-1}{j-1} \binom{D}{m} \binom{T}{1} \int_0^{\tau} [F(t)]^{j-1} [1-F(t)]^{T-j} \cdot [G(t)]^m [1-G(t)]^{D-m} \cdot f(t) dt \\
&= \binom{T-1}{j-1} \binom{D}{m} \binom{T}{1} \int_0^{\tau} [1-\exp(-v_t t)]^{j-1} [\exp(-v_t t)]^{T-j} \cdot [1-\exp(-v_d t)]^m [\exp(-v_d t)]^{D-m} \cdot v_t \exp(-v_t t) dt \\
&= v_t \binom{T-1}{j-1} \binom{D}{K-j} \binom{T}{1} \int_0^{\tau} [1-\exp(-v_t t)]^{j-1} [\exp(-v_t t)]^{T-j} \cdot [1-\exp(-v_d t)]^{K-j} [\exp(-v_d t)]^{D-K+j} \cdot \exp(-v_t t) dt \\
&= v_t \binom{T-1}{j-1} \binom{D}{K-j} \binom{T}{1} \int_0^{\tau} [1-\exp(-v_t t)]^{j-1} [\exp(-v_t t \cdot \{T-j\})] \cdot [1-\exp(-v_d t)]^{K-j} [\exp(-v_d t \cdot \{D-K+j\})] \cdot \exp(-v_t t) dt \\
&= v_t \binom{T-1}{j-1} \binom{D}{K-j} \binom{T}{1} \int_0^{\tau} [1-\exp(-v_t t)]^{j-1} [\exp(-v_t t \cdot \{T-j+1\})] \cdot [1-\exp(-v_d t)]^{K-j} [\exp(-v_d t \cdot \{D-K+j\})] dt \\
&= v_t \binom{T-1}{j-1} \binom{D}{K-j} \binom{T}{1} \int_0^{\tau} [\exp(-v_t t \cdot \{T-j+1\})] [\exp(-v_d t \cdot \{D-K+j\})] \cdot [1-\exp(-v_t t)]^{j-1} [1-\exp(-v_d t)]^{K-j} dt \\
&= v_t \binom{T-1}{j-1} \binom{D}{K-j} \binom{T}{1} \int_0^{\tau} [\exp(-v_t t \cdot \{T-j+1\} - v_d t \cdot \{D-K+j\})] \cdot \sum_{a=0}^{j-1} \left[\binom{j-1}{a} \cdot \{-\exp(-v_t t)\}^a \right] \sum_{b=0}^{K-j} \left[\binom{K-j}{b} \cdot \{-\exp(-v_d t)\}^b \right] dt \\
&= v_t \binom{T-1}{j-1} \binom{D}{K-j} \binom{T}{1} \int_0^{\tau} [\exp(-v_t t \cdot \{T-j+1\} - v_d t \cdot \{D-K+j\})] \cdot \sum_{a=0}^{j-1} \left\{ \binom{j-1}{a} \cdot \{-\exp(-v_t t)\}^a \right\} \cdot \sum_{b=0}^{K-j} \left[\binom{K-j}{b} \cdot \{-\exp(-v_d t)\}^b \right] dt
\end{aligned}$$

$$\begin{aligned}
&= v_t \binom{T-1}{j-1} \binom{D}{K-j} \binom{T}{1} \int_0^\tau [\exp(-v_t t \cdot \{T-j+1\} - v_d t \cdot \{D-K+j\})] \cdot \sum_{a=0}^{j-1} \left\{ \binom{j-1}{a} \cdot \sum_{b=0}^{K-j} \left[\binom{K-j}{b} \cdot \{-\exp(-v_t t)\}^a \cdot \{-\exp(-v_d t)\}^b \right] \right\} dt \\
&= v_t \binom{T-1}{j-1} \binom{D}{K-j} \binom{T}{1} \int_0^\tau \sum_{a=0}^{j-1} \left\{ \binom{j-1}{a} \cdot \sum_{b=0}^{K-j} \left[\binom{K-j}{b} \cdot [\exp(-v_t t \cdot \{T-j+1\} - v_d t \cdot \{D-K+j\})] \cdot \{-\exp(-v_t t)\}^a \cdot \{-\exp(-v_d t)\}^b \right] \right\} dt \\
&= v_t \binom{T-1}{j-1} \binom{D}{K-j} \binom{T}{1} \int_0^\tau \sum_{a=0}^{j-1} \left\{ \binom{j-1}{a} \cdot \sum_{b=0}^{K-j} \left[\binom{K-j}{b} \cdot [\exp(-v_t t \cdot \{T-j+1\} - v_d t \cdot \{D-K+j\})] \cdot (-1)^{a+b} \cdot \exp(-av_t t) \cdot \exp(-bv_d t) \right] \right\} dt \\
&= v_t \binom{T-1}{j-1} \binom{D}{K-j} \binom{T}{1} \int_0^\tau \sum_{a=0}^{j-1} \left\{ \binom{j-1}{a} \cdot \sum_{b=0}^{K-j} \left[\binom{K-j}{b} \cdot (-1)^{a+b} \cdot [\exp(-v_t t \cdot \{T-j+1\} - v_d t \cdot \{D-K+j\})] \cdot \exp(-av_t t - bv_d t) \right] \right\} dt \\
&= v_t \binom{T-1}{j-1} \binom{D}{K-j} \binom{T}{1} \sum_{a=0}^{j-1} \left\{ \binom{j-1}{a} \cdot \sum_{b=0}^{K-j} \left[\binom{K-j}{b} \cdot (-1)^{a+b} \int_0^\tau [\exp(-v_t t \cdot \{T-j+1\} - v_d t \cdot \{D-K+j\})] \cdot \exp(-av_t t - bv_d t) dt \right] \right\} \\
&= v_t \binom{T-1}{j-1} \binom{D}{K-j} \binom{T}{1} \sum_{a=0}^{j-1} \left\{ \binom{j-1}{a} \cdot \sum_{b=0}^{K-j} \left[\binom{K-j}{b} \cdot (-1)^{a+b} \int_0^\tau [\exp(-v_t t \cdot \{T-j+1\} - v_d t \cdot \{D-K+j\} - av_t t - bv_d t) dt] \right] \right\} \\
&= v_t \binom{T-1}{j-1} \binom{D}{K-j} \binom{T}{1} \sum_{a=0}^{j-1} \left\{ \binom{j-1}{a} \cdot \sum_{b=0}^{K-j} \left[\binom{K-j}{b} \cdot (-1)^{a+b} \int_0^\tau [\exp(-v_t t \cdot \{T-j+1+a\} - v_d t \cdot \{D-K+j+b\})] dt \right] \right\} \\
&= v_t \binom{T-1}{j-1} \binom{D}{K-j} \binom{T}{1} \sum_{a=0}^{j-1} \left\{ \binom{j-1}{a} \cdot \sum_{b=0}^{K-j} \left[\binom{K-j}{b} \cdot (-1)^{a+b} \int_0^\tau [\exp([-v_t \{T-j+1+a\} - v_d \{D-K+j+b\}] \cdot t)] dt \right] \right\} \\
&= v_t \binom{T-1}{j-1} \binom{D}{K-j} \binom{T}{1} \sum_{a=0}^{j-1} \left\{ \binom{j-1}{a} \cdot \sum_{b=0}^{K-j} \left[\binom{K-j}{b} \cdot \frac{(-1)^{a+b}}{v_t \{T-j+1+a\} + v_d \{D-K+j+b\}} [1 - \exp([-v_t \{T-j+1+a\} - v_d \{D-K+j+b\}] \cdot \tau)] \right] \right\}
\end{aligned}$$

After inserting v_t and v_d given by α , C , T , and D :

$$\begin{aligned}
P_2 &= v_t \binom{T-1}{j-1} \binom{D}{K-j} \binom{T}{1} \sum_{a=0}^{j-1} \left\{ \binom{j-1}{a} \cdot \sum_{b=0}^{K-j} \left[\binom{K-j}{b} \cdot \frac{(-1)^{a+b}}{v_t \{T-j+1+a\} + \alpha v_t \{D-K+j+b\}} \left[1 - \exp([-v_t \{T-j+1+a\} - \alpha v_t \{D-K+j+b\}] \cdot \tau) \right] \right] \right\} \\
&= v_t \binom{T-1}{j-1} \binom{D}{K-j} \binom{T}{1} \sum_{a=0}^{j-1} \left\{ \binom{j-1}{a} \cdot \sum_{b=0}^{K-j} \left[\binom{K-j}{b} \cdot \frac{(-1)^{a+b}}{v_t \{T-j+1+a+\alpha(D-K+j+b)\}} \left[1 - \exp([-v_t \tau \{T-j+1+a+\alpha(D-K+j+b)\}] \cdot 1) \right] \right] \right\} \\
&= \binom{T-1}{j-1} \binom{D}{K-j} \binom{T}{1} \sum_{a=0}^{j-1} \left\{ \binom{j-1}{a} \cdot \sum_{b=0}^{K-j} \left[\binom{K-j}{b} \cdot \frac{(-1)^{a+b}}{T-j+1+a+\alpha(D-K+j+b)} \left[1 - \exp\left[-\frac{C\tau}{T+\alpha D} \{T-j+1+a+\alpha(D-K+j+b)\}\right] \right] \right] \right\}
\end{aligned}$$

P_3 is the probability that the score equals j and the total number of elements equals K and the K -th element entering the VSTM is a *distractor*. The number of distractors entering VSTM denoted by m is always $K-j$. If $j=K$, or $j<K-D$, $P_3=0$; otherwise:

$$\begin{aligned}
P_3 &= \int_0^\tau \binom{T}{j} [F(t)]^j [1-F(t)]^{T-j} \cdot \binom{D-1}{m-1} [G(t)]^{m-1} [1-G(t)]^{D-m} \binom{D}{1} \cdot g(t) dt \\
&= \binom{T}{j} \binom{D-1}{m-1} \binom{D}{1} \int_0^\tau [F(t)]^j [1-F(t)]^{T-j} \cdot [G(t)]^{m-1} [1-G(t)]^{D-m} \cdot g(t) dt \\
&= \binom{T}{j} \binom{D-1}{K-j-1} \binom{D}{1} \int_0^\tau [1-\exp(-v_t t)]^j [\exp(-v_t t)]^{T-j} \cdot [1-\exp(-v_d t)]^{K-j-1} [\exp(-v_d t)]^{D-K+j} \cdot v_d \exp(-v_d t) dt \\
&= v_d \binom{T}{j} \binom{D-1}{K-j-1} \binom{D}{1} \int_0^\tau [1-\exp(-v_t t)]^j [\exp(-v_t t \cdot \{T-j\})] \cdot [1-\exp(-v_d t)]^{K-j-1} [\exp(-v_d t \cdot \{D-K+j\})] \cdot \exp(-v_d t) dt \\
&= v_d \binom{T}{j} \binom{D-1}{K-j-1} \binom{D}{1} \int_0^\tau [\exp(-v_t t \cdot \{T-j\} - v_d t \cdot \{D-K+j+1\})] \cdot [1-\exp(-v_t t)]^j [1-\exp(-v_d t)]^{K-j-1} dt \\
&= v_d \binom{T}{j} \binom{D-1}{K-j-1} \binom{D}{1} \int_0^\tau [\exp(-v_t t \cdot \{T-j\} - v_d t \cdot \{D-K+j+1\})] \cdot \sum_{a=0}^j \left[\binom{j}{a} \{-\exp(-v_t t)\}^a \right] \cdot \sum_{b=0}^{K-j-1} \left[\binom{K-j-1}{b} \{-\exp(-v_d t)\}^b \right] dt \\
&= v_d \binom{T}{j} \binom{D-1}{K-j-1} \binom{D}{1} \int_0^\tau [\exp(-v_t t \cdot \{T-j\} - v_d t \cdot \{D-K+j+1\})] \cdot \sum_{a=0}^j \left\{ \binom{j}{a} \{-\exp(-v_t t)\}^a \cdot \sum_{b=0}^{K-j-1} \left[\binom{K-j-1}{b} \{-\exp(-v_d t)\}^b \right] \right\} dt \\
&= v_d \binom{T}{j} \binom{D-1}{K-j-1} \binom{D}{1} \int_0^\tau [\exp(-v_t t \cdot \{T-j\} - v_d t \cdot \{D-K+j+1\})] \cdot \sum_{a=0}^j \left\{ \binom{j}{a} \cdot \sum_{b=0}^{K-j-1} \left[\binom{K-j-1}{b} \cdot \{-\exp(-v_t t)\}^a \cdot \{-\exp(-v_d t)\}^b \right] \right\} dt
\end{aligned}$$

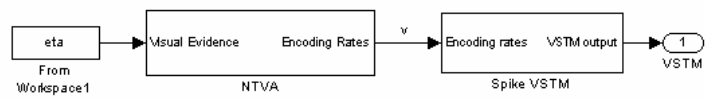
$$\begin{aligned}
&= v_d \binom{T}{j} \binom{D-1}{K-j-1} \binom{D}{1} \int_0^\tau \sum_{a=0}^j \left\{ \binom{j}{a} \cdot \sum_{b=0}^{K-j-1} \left[\binom{K-j-1}{b} \cdot (-1)^{a+b} \cdot \exp(-v_t t \cdot \{T-j\} - v_d t \cdot \{D-K+j+1\}) \cdot \exp(-av_t t - bv_d t) \right] \right\} dt \\
&= v_d \binom{T}{j} \binom{D-1}{K-j-1} \binom{D}{1} \sum_{a=0}^j \left\{ \binom{j}{a} \cdot \sum_{b=0}^{K-j-1} \left[\binom{K-j-1}{b} \cdot (-1)^{a+b} \cdot \int_0^\tau \exp(-v_t t \cdot \{T-j\} - v_d t \cdot \{D-K+j+1\} - av_t t - bv_d t) dt \right] \right\} \\
&= v_d \binom{T}{j} \binom{D-1}{K-j-1} \binom{D}{1} \sum_{a=0}^j \left\{ \binom{j}{a} \cdot \sum_{b=0}^{K-j-1} \left[\binom{K-j-1}{b} \cdot (-1)^{a+b} \cdot \int_0^\tau \exp(-v_t t \cdot \{T-j+a\} - v_d t \cdot \{D-K+j+1+b\}) dt \right] \right\} \\
&= v_d \binom{T}{j} \binom{D-1}{K-j-1} \binom{D}{1} \sum_{a=0}^j \left\{ \binom{j}{a} \cdot \sum_{b=0}^{K-j-1} \left[\binom{K-j-1}{b} \cdot (-1)^{a+b} \cdot \int_0^\tau \exp([-v_t \cdot \{T-j+a\} - v_d \cdot \{D-K+j+1+b\}] \cdot t) dt \right] \right\} \\
&= v_d \binom{T}{j} \binom{D-1}{K-j-1} \binom{D}{1} \sum_{a=0}^j \left\{ \binom{j}{a} \cdot \sum_{b=0}^{K-j-1} \left[\binom{K-j-1}{b} \cdot \frac{(-1)^{a+b}}{v_t \cdot \{T-j+a\} + v_d \cdot \{D-K+j+1+b\}} \cdot [1 - \exp([-v_t \cdot \{T-j+a\} - v_d \cdot \{D-K+j+1+b\}] \cdot \tau)] \right] \right\}
\end{aligned}$$

After inserting v_t and v_d given by α , C , T , and D :

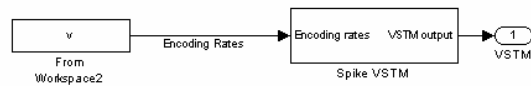
$$\begin{aligned}
P_3 &= \alpha v_t \binom{T}{j} \binom{D-1}{K-j-1} \binom{D}{1} \sum_{a=0}^j \left\{ \binom{j}{a} \cdot \sum_{b=0}^{K-j-1} \left[\binom{K-j-1}{b} \cdot \frac{(-1)^{a+b}}{v_t \cdot \{T-j+a\} + \alpha v_t \cdot \{D-K+j+1+b\}} \cdot [1 - \exp([-v_t \cdot \{T-j+a\} - \alpha v_t \cdot \{D-K+j+1+b\}] \cdot \tau)] \right] \right\} \\
&= \alpha \binom{T}{j} \binom{D-1}{K-j-1} \binom{D}{1} \sum_{a=0}^j \left\{ \binom{j}{a} \cdot \sum_{b=0}^{K-j-1} \left[\binom{K-j-1}{b} \cdot \frac{(-1)^{a+b}}{T-j+a + \alpha \{D-K+j+1+b\}} \cdot [1 - \exp([-v_t \tau \cdot \{T-j+a + \alpha \cdot (D-K+j+1+b)\}])] \right] \right\} \\
&= \alpha \binom{T}{j} \binom{D-1}{K-j-1} \binom{D}{1} \sum_{a=0}^j \left\{ \binom{j}{a} \cdot \sum_{b=0}^{K-j-1} \left[\binom{K-j-1}{b} \cdot \frac{(-1)^{a+b}}{T-j+a + \alpha \{D-K+j+1+b\}} \cdot \left[1 - \exp\left(-\frac{C\tau}{T+\alpha D} \cdot \{T-j+a + \alpha \cdot (D-K+j+1+b)\}\right) \right] \right] \right\}
\end{aligned}$$

Appendix B Simulink diagrams

The GrandScheme, ver. 1 (main diagram)

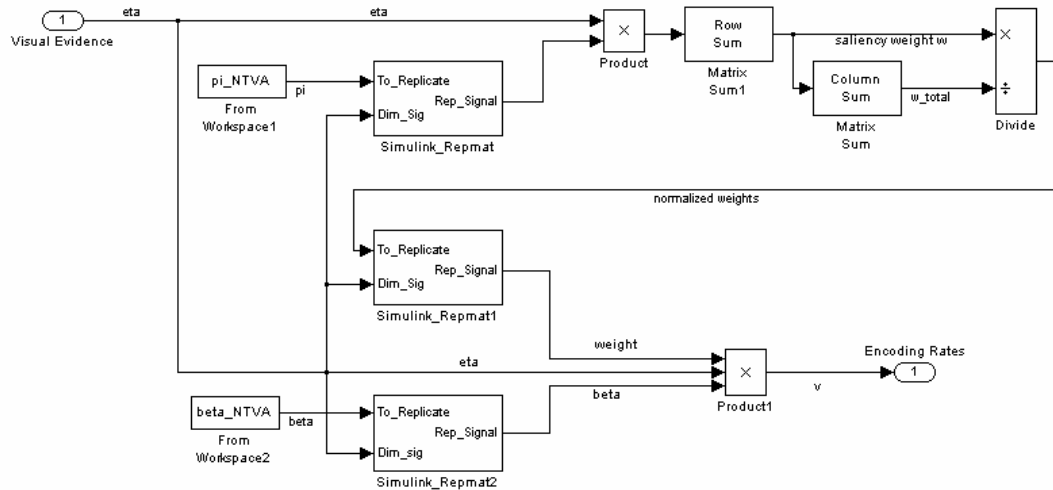


The GrandScheme, ver. 2 (main diagram – simple version)

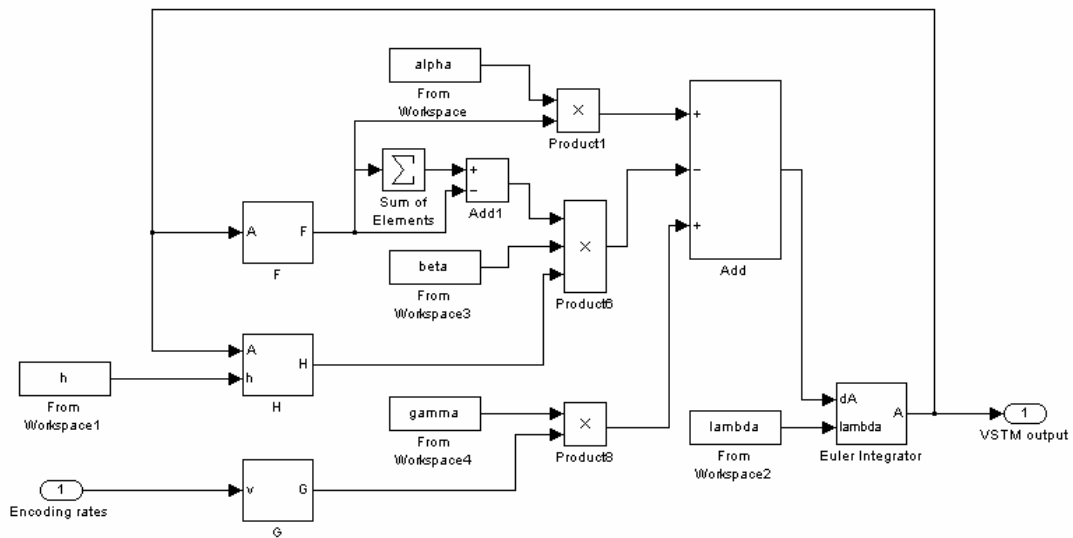


The NTVA Block (child of GrandScheme)

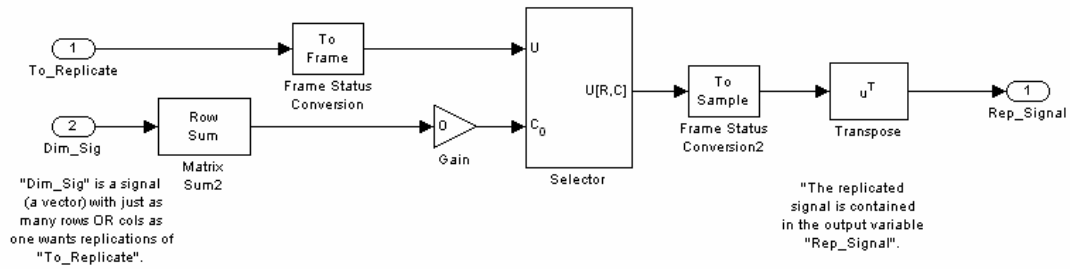
NBI ONE CAN USE
A MATRIX VIEWER
FROM THE SIGNAL
PROCESSING
SIMULINK TOOLBOX
TO VIEW THE MATRIX
SIGNALS



The Spike VSTM block (child of GrandScheme)

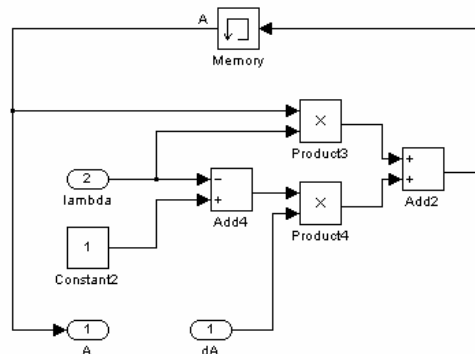


The repmat block (child of NTVA, child of GrandScheme)

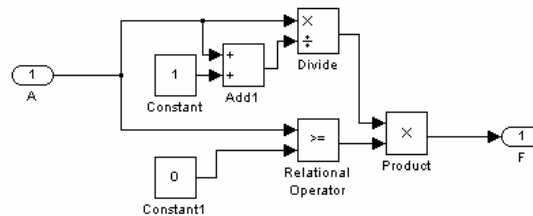


NB! To make this sub system open at start-up choose:
 (right-click)
 -mask subsystem
 -invisible/visible
 -apply, save
 -reopen main model

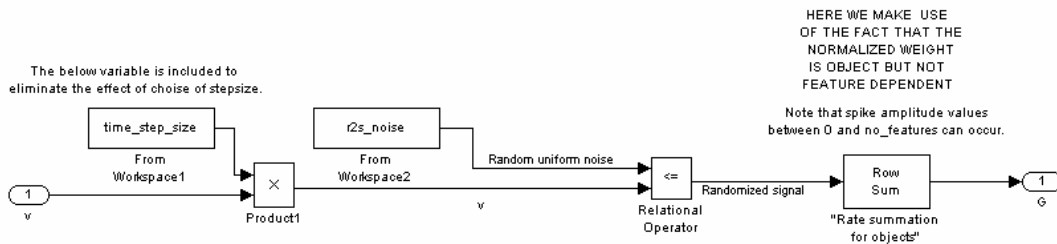
The Euler Integrator block (child of Spike VSTM, child of GrandScheme)



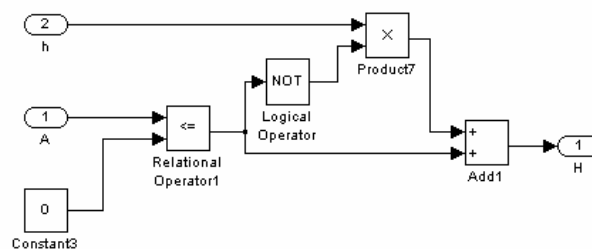
The F block (child of Spike VSTM, child of GrandScheme)



The G block (child of Spike VSTM, child of GrandScheme)



The H block (child of Spike VSTM, child of GrandScheme)



Appendix C Paper for NIPS2006

Visual Attention – A Neural Network Model of the Visual Short-Term Memory

Anders Petersen ^a, Søren Kyllingsbæk ^b & Lars Kai Hansen ^a

^a Informatics and Mathematical Modeling, Technical University of Denmark,
Richard Petersens Plads, B. 321, 2800 - Kgs. Lyngby, Denmark

^b Department of Psychology, Copenhagen University,
Linnésgade 22, 1361 - Copenhagen K, Denmark

Abstract

In this paper a neural network model of Visual Short-Term Memory (VSTM) is presented. The model links closely with Bundesen's (1990) well-established mathematical theory of visual attention. We evaluate the model's ability to fit experimental data from a classical whole and partial report study. Previous statistic models have successfully assessed the spatial distribution of visual attention; our neural network meets this standard and offers a neural interpretation of how objects are consolidated in VSTM at the same time. We hope that in the future, the model will be able to fit temporally dependent phenomena like the attentional blink effect, lag-1 sparing, and attentional dwell-time.

1 Introduction

In everyday life, it is important for us to be able to perceive, comprehend, and react to events in our environment. Often, our rate of success is heavily dependent upon how efficient and how fast we can process, interpret and react to sensory stimuli, e.g. like when we are driving a car.

In the following we shall refer to *visual attention* as the process that enables us to focus our processing resources to certain important objects in the visual scene. Following the Theory of Visual Attention [3] we assume that features have already been extracted and objects successfully segregated on the basis of their individual feature spaces. Our model deals with the important question of how only a limited sub span of all objects are actually selected and further encoded into VSTM.

Cartell already in the late 19th century demonstrated a surprising limit in how many objects that can be perceived at the same time – a limit only about 4 objects which may be held in the VSTM at the same time [4],[5]. This finding is independent of the number of objects visually presented at the same time [11]. Evidence further exist that the “magical number” of 3-to-4 objects is largely independent of how many features are encoded for each object, i.e. the complexity of the visual object, does not hold an influence on the memorial capacity of the VSTM (see [8], but see also [1]).

Modelling the function of the VSTM, it is essential that the inherent capacity limitation is properly mimicked, since it seems a fundamental limit of the system. Most likely the VSTM would be heavily overloaded, should the system lack the ability to represent only the most salient of the visually appearing objects.

2 The Model

The model that we are presenting in this paper can actually be understood as three important consecutive processes.

The first process is simply extraction of visual features, we speak of this process as *'object matching'*, since we find it relevant to think that objects in the visual field are to some extent *'matched'* against objects representations in Visual Long-Term Memory (VLTM). In this paper we do not consider the problem of which feature extraction techniques are biologically most plausible or perhaps technically most appropriate to use.

The second process that we shall consider in more detail is *'the attentional race'*. According to Shibuya & Bundesen (1988), all objects in the visual scene take a place in what one could think of as a race to become encoded. In Shibuya & Bundesen's race model, the *'odds'* that a given object is selected as a winner in the race is directly related to the rate value with which the object participates. It is worth noting that the race is a stochastic, rather than a deterministic process, meaning that no one can beforehand predict readily which objects will win the race.

The third and last process that we shall consider is that of *'storage'* of object representation in VSTM. Inspired by Usher & Cohen (1999) we propose a competitive neural network model of VSTM, directly linking with several important assumptions expressed in Bundesen's Theory of Visual Attention [3].

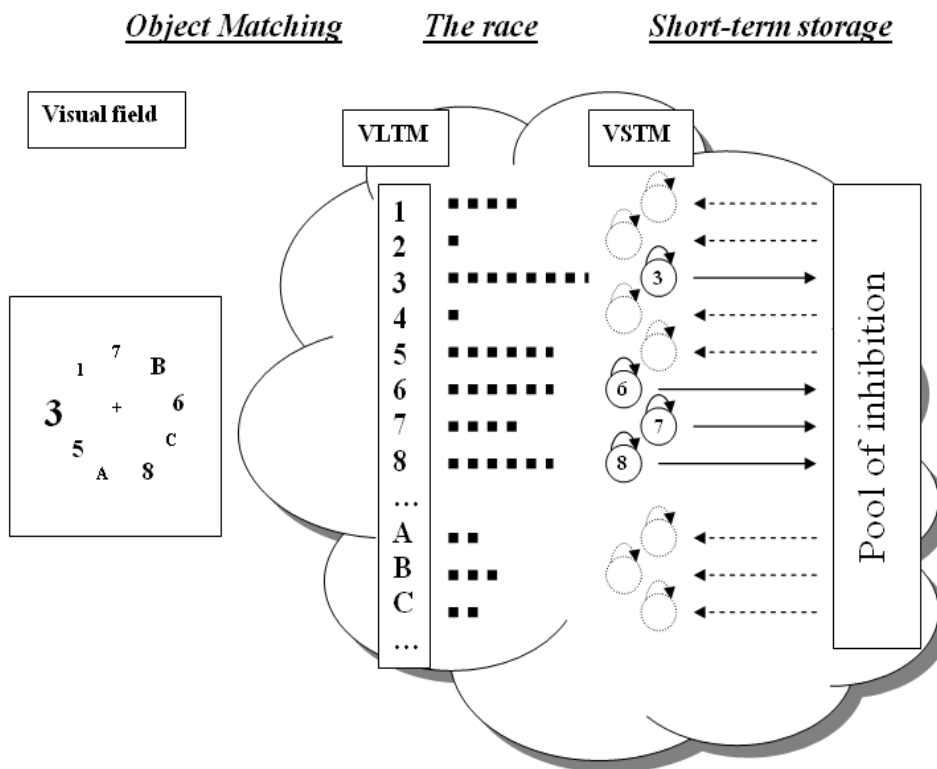


Figure 1: The model scheme – an example

An overview of the three processes mentioned above is given Figure 1. During a classical partial report experiment, the subject is presented with a visual field consisting of a number of digits and letters. The subject is asked to report digits only. We see that among the targets, those digits that occupy a spatially larger area are more likely to enter the VSTM.

In principle - and practice however also smaller digits can enter the VSTM, as can also the distractors, i.e. the letters. The figure depicts a typical situation where four objects have succeeded in consolidating themselves in VSTM.

2.1 The neural Theory of visual Attention

The theory of visual attention (TVA) proposed by Bundesen (1990) is a unified theory of visual recognition and attentional selection. TVA provides a mathematical framework describing how the visual system is able to select individual objects in the visual field, based on the visual evidence and the setting of two different types of visual preference parameters (pertinence and bias), representing the influence from higher cortical areas, including VLTM.

The output of the TVA-model is a set of rate parameters v that are directly related to the probability that a given characterization, *object x belongs to category i* , is encoded into the VSTM. The rate parameters are given by:

$$v(x, i) = \eta(x, i) \beta_i \frac{w_x}{\sum_{z \in S} w_z} \quad (1)$$

Where

$$w_x = \sum_{j \in R} \eta(x, j) \pi_j \quad (2)$$

Here $\eta(x, i)$ is defined as the strength of the sensory evidence that object x belongs to the visual category i . The pertinence of the visual category j is denoted by π_j and setting of these values effectively implements the so-called filtering mechanism. The perceptual decision bias of a visual category i is denoted by β_i and setting of these values conversely implements a complementary mechanism called pigeonholing.

The filtering mechanism increases the likelihood that elements belonging to a target category are perceived, without biasing perception in favor of perceiving the elements as belonging to any particular category.

Pigeonholing, conversely changes the probability that a particular category i is selected without affecting the conditional probability that element x is selected given that category i is selected.

A neural interpretation of TVA is given in [2]. Basically here pigeonholing (selection of features) is considered an increase in the rate of firing of neurons while filtering (selection of objects) is considered an increased mobilization of neurons.

Corresponding to the interpretation in NTVA the fraction $w_x / \sum w_z$ in equation (1), which is the relative attentional weight of object x compared to the weight of all objects z in the visual field S , can be directly interpreted as the relative fraction of neurons allocated to process a given object x , compared to the total number of neurons processing just any object z belonging to the visual field S .

Each and every characterization generally takes the form *object x belongs to category i* .

Denoting the set of all features as R the total processing capacity, can be considered a constant C , which equals the sum of all encoding rates v (See [3]).

$$C = \sum_{x \in S} \sum_{i \in R} v(x, i) \quad (3)$$

Shibuya and Bundesen (1988) assume that rates of encoding for targets, v_T and for distractors, v_D can be calculated according to the formulas:

$$v_T = \frac{C}{T + \alpha D} \quad v_D = \frac{\alpha C}{T + \alpha D} = \alpha v_T \quad (4)$$

Here α characterizes the ratio of discrimination between distractors and targets.

The effective exposure duration τ is smaller than the actual exposure duration t by an amount t_0 corresponding to the temporal threshold before conscious processing begins. However the effective exposure duration can not be negative so computationally it is set to:

$$\tau = \max(0, t - t_0) \quad (5)$$

In our model we adopt the parameters C , α and t_0 and further, following Bundesen, we make use of equations (4)-(6).

2.2 The Neural Network model of VSTM

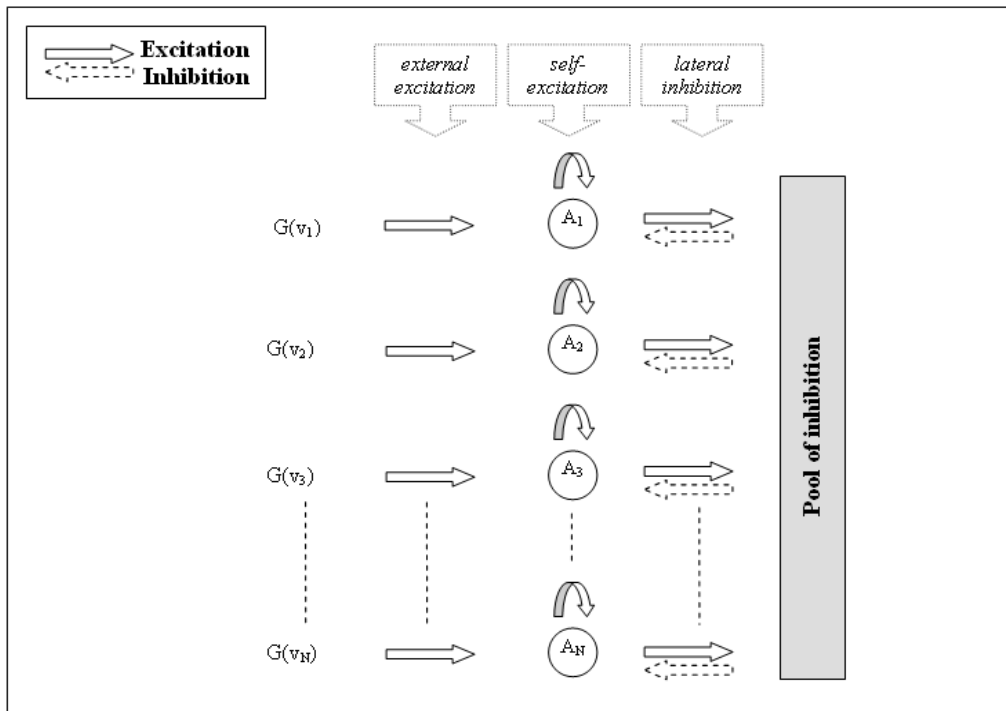


Figure 2: The Neural Network model of VSTM. The total number of neuron assemblies is N and each assembly is represented by a level of activation A .

An object can enter VSTM once it receives external excitation, G taking the shape of Poisson distributed spike trains, arriving with the rate parameter v . (See Figure 2).

A neural assembly that has obtained a positive level of activation will automatically seek to re-excite itself, so that it can stay in VSTM, at the same time trying to inhibit activation in other neuron assemblies representing other objects, i.e. working to suppress other object from co-temporally being stored in VSTM.

The initial condition for the simulations is that all neuron assemblies start with an activation of zero, i.e. no objects are initially stored in VSTM. As a consequence neither re-excitation nor lateral inhibition exists, before the assemblies are externally activated.

3 Implementation

The activation A_x of neuron assembly x (representing object x) is given by the first order differential equation:

$$\frac{dA_x}{dt} = -A_x + \alpha^* F(A_x) - \beta^* \sum_{z \neq x} F(A_z) + \gamma^* G(v_x) \quad (6)$$

The above equation characterizes a leaky accumulator model. There is passive decay of the activation towards the rest level, with a time constant chosen as 1, reflecting the time scale that physiologically is observed with synaptic currents [12].

F is a squashing function that keeps the activation within bounds:

$$\begin{aligned} F(A) &= 0, & \text{for } A \leq 0 \\ F(A) &= \frac{A}{1+A}, & \text{for } A > 0 \end{aligned} \quad (7)$$

As a consequence of the squashing function F , the parameter α^* is the limiting value of maximal self-excitation that assemblies can up-hold and the parameter β^* is the limiting maximal value of inhibition that can be sent from one assembly to another.

Also the model assume we can not have negative self-excitation, i.e. self-inhibition and further the model does not implement any terms that could account for excitation laterally between the assemblies. The latter effect could for instance be included if one wanted to account for semantically related objects and their effect on the number of reported objects.

The attentional significance that object i is present in the visual field R is represented by the encoding rate v_i . In our model we follow the approach from [3] and interpret this rate as the firing rate of a Poisson spike generator G . Hence γ^* characterizes the amplitude of the Poisson distributed input spikes arriving to the neuron assembly x .

The model was implemented in MATLAB's Simulink toolbox. At least in the operated parameter domain we judge the stiffness of the system to be negligible so for simplicity we numerically apply Euler integration with integration step size $dt = 0.01$.¹

4 Model performance

4.1 The data set

The data covers the performance of a single subject, participating in an extensive series of whole and partial report experiments. The subject was instructed to report targets, i.e. digits while ignoring distractors, i.e. letters displayed on an imaginary circle around a small fixation cross at the center of the screen. Experimental trials covered twelve different combinations of total number of 2 – 6 targets, T , and 0 – 6 distractors, D . Further, exposure durations t were varied systematically at 10, 20, 30, 40, 50, 70, 100, 150 and 200 ms. Each experimental condition was repeated 60 times but trials were mixed so that the subject had no a-priori knowledge of the experimental condition. Moreover trials were grouped into

¹ We verified that we used an appropriately small step size in our update formula in order not to consider the influence from having more than one spike per time interval, the probability for more than one spike was calculated to 0.37 %.

4.2 Performance of the Neural Network model

Figure 3 shows accumulated score distributions. The score is defined as the number of targets reported correctly. The upper most curve represents the accumulated score of $j = 1$, i.e. the probability of reporting 1 or more targets correctly. Other curves represent accumulated probabilities for reporting at least 2, 3, 4 or even 5 targets.

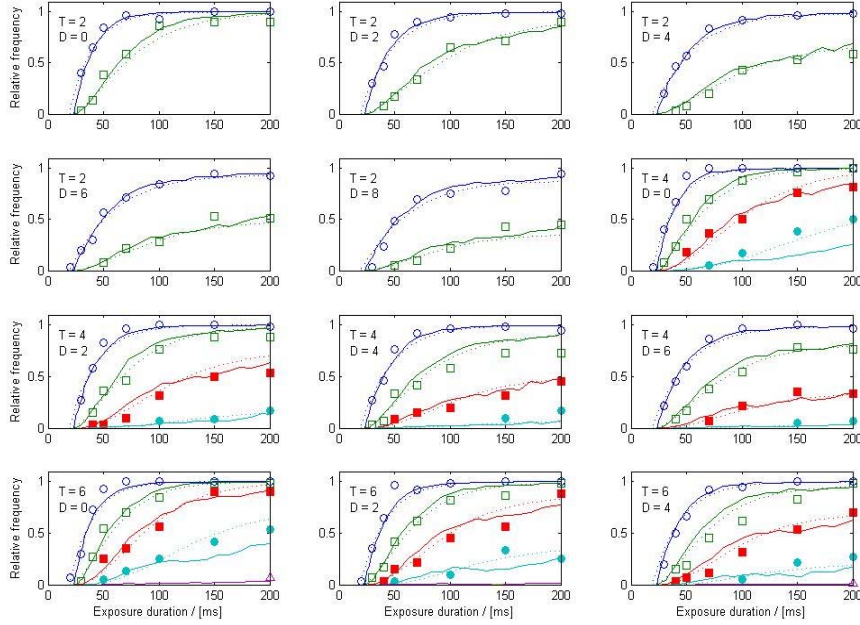


Figure 3: Accumulated score distribution for subject MP in [10]. Empirically found values are plotted with symbols as markers. The dotted lines represent the fit by Shibuya & Bundesen in [10]. Solid lines represent the performance of our neural network model.

Shibuya and Bundesen (1988) proposed a mixture model, mixing probabilities obtained with using a statistical model that assumed memorial capacities of either $K = 3$ or $K = 4$ respectively.

There is a relatively close fit between the proposed mixture model and the empirical data. We see however that data points obtained with exposure duration around 50 ms are generally under fitted and more noticeably the model does not account for cases where more than 4 targets are reported, as is actually the case in two out of three of the lower most plots.

What we observe with the previous model can be considered a trade-off between two conflicting demands. The first demand is to fit the initial part of the curves, i.e. the larger the processing capacity C the steeper the curves will rise, on the other hand the second demand, which is to keep the score distribution reasonably low for long exposure durations, require that the processing capacity C is not set too high. Hence the setting of C is set subject to a compromise.

Addressing the performance of our neural network model we think it clearly meets the standard of Shibuya and Bundesen's model. Moreover, and in contrast to Shibuya and Bundesen's model, our new model readily demonstrates its capability of predicting extreme cases, where more than 4 objects are reported.

5 Discussion

Our new dynamic model of visual attention and VSTM is able to account for the complete set of data from whole and partial report experiments. Where the previous account by Shibuya and Bundesen [10] treated extreme scores as outliers, the new model encompasses these as natural consequences of the internal dynamics. Further, the model explains VSTM capacity and consolidation as the result of a dynamic process rather than as a static store, which capacity is independent of processing capacity and the attentional set of the subject.

In future studies, we wish to explore the model's ability to explain the dynamic consolidation in VSTM found in temporally extended paradigms such as the attentional blink paradigm and studies of attentional dwell time (e.g. [13]). Here, consolidation in VSTM is strongly dependent on competition between items already encoded into VSTM and visual items presented at a later point in time. This competitive process follows naturally from the dynamic architecture of the present model.

Acknowledgements

We would like to thank Claus Bundesen for access to the experimental data in [10]. Also we would like to thank Marius Usher for kindly verifying parameters and settings in [12].

References

- [1] Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science, 15*, 106-111.
- [2] Bundesen, C., Habekost, T. & Kyllingsbæk, S. (2005). A Neural Theory of Visual Attention: Bridging Cognition and Neurophysiology. *Psychological Review, 112*, 291-328.
- [3] Bundesen, C. (1990). A Theory of visual Attention. *Psychological Review, 97*, 523-547.
- [4] Cattell, J. M. (1886). The inertia of the eye and brain. *Brain, 8*, 295-312.
- [5] Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87-114.
- [6] Davelaar, E. J., Haarmann, H. J., Goshen-Gottstein, Y., Ashkenazi, A. & Usher, M. (2005). The Demise of Short-Term Memory Revisited: Empirical and Computational Investigations of Recency Effects. *Psychological Review, 112*, 3-42.
- [7] Davelaar, E. J. & Usher, M. (2004). An Extended Buffer Model for Active Maintenance and Selective Updating. *Connectionist Models of Cognition and Perception II: Proceedings of the 8th Neural Computation and Psychology Workshop*, 3-14.
- [8] Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*, 279-281.
- [9] Nieuwenhuis, S., Holmes, B. D., Gilzenrat, M. S. & Cohen, J. D. (2005). The role of the Locus Coeruleus in Mediating the Attentional Blink: A Neurocomputational Theory. *Journal of Experimental Psychology: General, 134*, 291-307.
- [10] Shibuya, H. & Bundesen, C. (1988). Visual Selection From Multielement Displays: Measuring and Modeling Effects of Exposure Duration. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 591-600.
- [11] Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied, 74*(11), 1-30.
- [12] Usher, M. & Cohen, J. D. (1999). Short Term Memory and Selection Processes in a Frontal-Lobe Model. *Connectionist Models in Cognitive Neuroscience, 78-91*.
- [13] Ward, R., Duncan, J. & Shapiro, K. (1996). The slow Time-Course of Visual Attention. *Cognitive Psychology, 30*, 79-109.

Nomenclature

Abbreviations

Abbreviation	Explanation
2CMM	2-Component Mixture Model
2CMM ₂	Corrected 2-Component Mixture Model
AIC	Akaike Information Criterion
BIC	Bayes Information Criterion
CNUSM	Conservatory Non-Unit Spike Model
CVC	Center for Visual Cognition
ECTS	European Credit Transfer System
FIRM	Fixed-capacity Independent Race Model
FP	Fokker-Planck
ICA	Independent Component Analysis
IMM	Informatics and Mathematical Modelling
ISI	Inter-Stimulus Interval
MCMM	Multi-Component Mixture Model
ML	Maximum Likelihood
NIPS	Neural Information Processing Systems
NTVA	Neural Theory of Visual Attention
NUSM	Non-Unit Spike Model
NLL	Negative log-likelihood
PR	Partial Report
RAM	Random Access Memory
RSVP	Rapid Serial Visual Presentation
SOA	Stimulus Onset Asynchrony
SDE	Stochastic Differential Equation
STM	Short-Term Memory
USM	Unit Spike Model
TVA	Theory of Visual Attention
VLTM	Visual Long-Term Memory
VSTM	Visual Short-Term Memory
WR	Whole Report

Symbols

Symbol	Explanation
A	Neuron assembly activation level
A_{th}	Activation threshold
C	Processing capacity of the VSTM
D	Number of distractors
h	Inhibition discrimination factor
I	Neuronal input
j	The number of reported targets
K	Storage capacity of the VSTM
L	Likelihood
N_{na}	Number of neuron assemblies
N_{fp}	Number of free parameters
R	Set of visual features
s	Discrete time step
S	Visual field
t	Time
t_e	Exposure duration
t_0	Temporal threshold for conscious processing
T	Number of targets
v	Rate of processing
w	Attentional weight
α	Visual discrimination factor
α^*	Level of self-excitation
β	Perceptual decision bias
β^*	Level of lateral inhibition
γ	Spike amplitude of Poisson input
η	Visual evidence
θ	Experimental condition
λ	Euler integration constant
π	Pertinence
τ	Effective exposure duration

References

- [1] Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*, 106-111.
- [2] Binomial distribution. (2006, July 26). In *Wikipedia, The Free Encyclopedia*. Retrieved 10:10, July 27, 2006, from http://en.wikipedia.org/w/index.php?title=Binomial_distribution&oldid=65993412.
- [3] Bundesen, C., Habekost, T., & Kyllingsbæk, S. (2005). A Neural Theory of Visual Attention: Bridging Cognition and Neurophysiology. *Psychological Review*, *112*, 291-328.
- [4] Bundesen, C. & Harms, L. (1999). Single-letter recognition as a function of exposure duration. *Psychological Research*, *62*, 275-279.
- [5] Bundesen, C. (1990). A Theory of visual Attention. *Psychological Review*, *97*, 523-547.
- [6] Bundesen, C. (1987). Visual Attention: Race model for selection from multielement displays. *Psychological Research*, *49*, 113-121.
- [7] Cattell, J. M. (1886). The inertia of the eye and brain. *Brain*, *8*, 295-312.
- [8] Chun, M. M. & Potter, M. C. (1995). A Two-Stage Model for Multiple Target Detection in Rapid Serial Visual Presentation. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 109-127.
- [9] Cotterill, Rodney (1998). *Enchanted looms: conscious networks in brains and computers*. Cambridge, United Kingdom. Cambridge University Press, (ISBN 0-521-62435-5).
- [10] Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioural and Brain Sciences*, *24*, 87-114.
- [11] Davelaar, E. J., Haarmann, H. J., Goshen-Gottstein, Y., Ashkenazi, A. & Usher, M. (2005). The Demise of Short-Term Memory Revisited: Empirical and Computational Investigations of Recency Effects. *Psychological Review*, *112*, 3-42.
- [12] Davelaar, E. J. & Usher, M. (2004). An Extended Buffer Model for Active Maintenance and Selective Updating. *Connectionist Models of Cognition and Perception II: Proceedings of the 8th Neural Computation and Psychology Workshop*, 3-14.
- [13] Dayan, P. and Abbott, L. F. (2001). *Theoretical Neuroscience: computational and mathematical modeling of neural systems*. Cambridge, Massachusetts, The MIT Press (ISBN 0-262-04199-5).

-
- [14] Deco, G. (2005). The computational Neuroscience of Visual Cognition: Attention, Memory and Reward, *WAPCV 2004, LNCS 3368*, 100-117.
- [15] Deco, G. & Zihl, J. (2004). A biased competition based neurodynamical model of visual neglect. *Medical Engineering & Physics*, 26, 733-743.
- [16] Deco, G. & Edmund T. Rolls (2003). A Neurodynamical cortical model of visual attention and invariant object recognition, *Vision research*, 44, 621-642.
- [17] Deco, G. & Rolls, E. T. (2003). Attention and working memory: a dynamical model of neuronal activity in the prefrontal cortex. *European Journal of Neuroscience*, 18, 2374-2390.
- [18] Deco, G. & Zihl, J. (2001). A Neurodynamical Model of Visual Attention: Feedback Enhancement of Spatial Resolution in a Hierarchical System. *Journal of Computational Neuroscience*, 10, 231-253.
- [19] Duncan, J., Ward, R. & Shapiro, K. (1994). Direct measurement of attentional dwell time in human vision. *Nature*, vol. 369.
- [20] Fragopanagos, N., Kockelkoren, S. & Taylor, J. G. (2005). A Neurodynamic model of the attentional blink. *Cognitive Brain Research*, 24, 568-586.
- [21] Hebb, D. O. (1949). *The organization of behaviour*, New York: Wiley.
- [22] Hypergeometric distribution. (2006, July 25). In *Wikipedia, The Free Encyclopedia*. Retrieved 10:03, July 27, 2006, from http://en.wikipedia.org/w/index.php?title=Hypergeometric_distribution&oldid=65801488.
- [23] Hyvärinen, A., Hoyer, P. O, Hurri, J. & Gutmann, M. (2005). Statistical Models of Images and Early Vision. *Proceedings of the Int. Symposium on Adaptive Knowledge Representation and Reasoning (AKRR2005)*, Espoo, Finland.
- [24] Kessler, K., Gross, J., Schmitz F. & Schnitzler A. (2006). Cortical dynamics and synchronization related to multiple target consolidation under rapid-serial-visual-presentation conditions. *Journal of Physiology – Paris*, 99, 21-28.
- [25] Kessler, K., Schmitz, F., Gross, J., Hommel, B. Shapiro, K. & Schnitzler, A. (2005). Target consolidation under high temporal processing demands as revealed by MEG. *NeuroImage*, 26, 1030-1041.
- [26] Kyllingsbæk, S. (2001). *Parallel and Serial Selective Processing in Vision*. Ph.D. dissertation, University of Copenhagen, Denmark <http://www.psy.ku.dk/cvc/SK/phd.htm>.
- [27] Lagarias, J.C., J. A. Reeds, M. H. Wright, & P. E. Wright. (1998). *Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions*, SIAM Journal of Optimization, Vol. 9 Number 1, pp. 112-147.
- [28] Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390, 279-281.

- [29] Miller, G. A. (1956), The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information. *Psychological Review*, 63, 81-97.
- [30] Nieuwenhuis, S., Holmes, B. D., Gilzenrat, M. S. & Cohen, J. D. (2005). The role of the Locus Coeruleus in Mediating the Attentional Blink: A Neurocomputational Theory. *Journal of Experimental Psychology: General*, 134, 291-307.
- [31] Di Lollo, V., Kawahara, J., Ghorashi, S. M. S. & Enns, J. T. (2005). The attentional blink: Resource depletion or temporary loss of control?, *Psychological Research*, 69, 191-200.
- [32] Moore's law. (2006, July 8). In *Wikipedia, The Free Encyclopedia*. Retrieved 09:57, July 9, 2006, from http://en.wikipedia.org/w/index.php?title=Moore%27s_law&oldid=62677016.
- [33] Phillips, W. A., & Baddeley, A. D. (1971). Reaction time and short-term visual memory. *Psychonomic Science*, 22(2), 73-74.
- [34] Potter, M. C. (2005). Competition for Attention in Space and Time: The First 200 ms. (*In press*).
- [35] Raffone, A. & Wolters, G. (2001). A Cortical Mechanism for binding in Visual Working Memory. *Journal of Cognitive Neuroscience*, 13, 766-785.
- [36] Usher, M., Olami, Z. & McClelland J. L. (2002). Hick's Law in a Stochastic Race Model with Speed-Accuracy Tradeoff. *Journal of Mathematical Psychology*, 1-12.
- [37] Usher, M. & Cohen, J. D. (1999). Short Term Memory and Selection Processes in a Frontal-Lobe Model. *Connectionist Models in Cognitive Neuroscience*, 78-91.
- [38] Shibuya, H. & Bundesen, C. (1988). Visual Selection From Multielement Displays: Measuring and Modeling Effects of Exposure Duration. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 591-600.
- [39] Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, 74(11), 1-30.
- [40] Ward, R., Duncan, J. & Shapiro, K. (1996). The slow Time-Course of Visual Attention. *Cognitive Psychology*, 30, 79-109.
- [41] Wrigley, S. N. & Brown, G. J. (2004). A Computational Model of Auditory Selective Attention, *IEEE Transactions on Neural Networks*, 15, 1151-1163.
- [42] Stochastic differential equation. (2006, June 26). In *Wikipedia, The Free Encyclopedia*. Retrieved 11:44, July 5, 2006, from http://en.wikipedia.org/w/index.php?title=Stochastic_differential_equation&oldid=60689702.
- [43] Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. Oxford, Oxford University Press Inc. (ISBN 0-19-853864-2).

