Modeling Continuous Representations in Visual Working Memory

Johannes Lohmann (johannes.lohmann@uni-tuebingen.de)
Martin V. Butz (martin.butz@uni-tuebingen.de)
Cognitive Modeling, Department of Computer Science, Department of Psychology,
Sand 14, Tübingen, 72076, Germany

Abstract

Visual working memory (VWM) is a crucial part of our cognitive system. Currently there is an active debate how the apparent limitations of VWM should be described. Limited-slot and flexible-resource theories are discussed, but so far the temporal dynamics of representations stored in VWM are not fully understood. In this paper we present data that supports the notion of dynamic VWM contents with changing precision. To account for these observations in a qualitative way, we propose a neural network that is able to account for emerging capacity limits as well as for changes in the precision of stored information.

Keywords: VWM; Visual Attention; Neural Network

Introduction

Successful interaction with our environment requires the storage and maintenance of task-relevant information. Planning of behavior as well as evaluation of the outcome would not be possible without this ability. The fact that we are able to perform these tasks indicates that there is some kind of interface between our environment and cognition. This interface is provided by the working memory system. Due to its importance in linking cognition to the external world, working memory has been studied extensively for decades.

An apparent limitation of working memory and especially visual working memory (VWM) with respect to the amount of preserved information has been observed throughout the years (see Miller, 1956 for an early review). Until today it is still in discussion how these limitations might be characterized. While *limited-slot* theories (Zhang & Luck, 2008) state that the maximum number of stored representations is limited and cannot be increased by decreasing the precision of individual representations, *flexible-resource* theories (Bays & Hussain, 2008) assume a mnemonic resource that can be used to either store a large number of low-precision representations or a small number of high-precision representations. The differentiation between both theories is fundamental as the alternatives suggest rather different bases of cognition.

There is empirical evidence for both alternatives (see Fukuda, Awh, & Vogel, 2010 for a recent review), even if resource models seem more plausible from a modeling perspective (Berg, Shin, Chou, George, & Ma, 2012). Recent studies investigated the neural basis for capacity limits, but again the results are mixed. While Anderson, Vogel, and Awh (2011) found behavioral as well as electrophysiological evidence for a limited number of slots in humans, Buschman, Siegel, Roy, and Miller (2011) reported mixed evidence in rhesus monkeys. The results imply the existence of discrete slots, containing some kind of resource that determines the precision of the stored representations.

To sum up, it remains unclear if capacity limits of VWM can be described in terms of slots or resources, but the empirical evidence indicates that it is unlikely that a strict resource or slot model will be the consensus – some kind of mixed model seems most probable. Interestingly, in contrast to the limitations with respect to the number of stored items, the temporal dynamics of VWM contents are far less investigated. Also in this respect the results are inconclusive. Zhang and Luck (2009) reported clear evidence for abrupt loss of information, whereas Salmela, Lahde, and Saarinen (2012) found evidence for the gradual loss of information. The results of Zhang and Luck (2009) are more in line with a slot model. The gradual decay observed by Salmela et al. (2012) is more in line with a resource model, in which the precision of the representations degrades over time. To investigate the temporal dynamics of information maintained in VWM we applied the same experimental paradigm as Zhang and Luck (2009). Furthermore, we developed a neural network model that can qualitatively account for the observations.

In the next section we describe the experimental setup. Next we report the obtained results. After this we give an outline of the neural model. A short discussion concludes the paper.

Experimental Setup

Since the influential study of Luck and Vogel (1997), change detection paradigms have become the standard approach in VWM research. As it was highlighted by Brady, Konkle, and Alvarez (2011), however, change detection paradigms do not allow to assess the precision of the representation that underlies the response of the participants. Zhang and Luck (2009) proposed a paradigm that allows to obtain a measure for the precision of the stored representations.

Participants had to remember Fourier descriptors which varied continuously in their phase (see Fig. 1 for the trial sequence and the stimuli). After the presentation of the stimuli a response screen appeared after a variable interstimulus interval (ISI). The position of one of the presented shapes was highlighted and participants had to indicate the presented stimuli on a "shape-wheel" containing the whole shape dimension. If the critical shape was in working memory, participants should report a shape close to the original shape, the response distribution should be bell-shaped and its deviation would be a measure of the precision of the representation. Without a representation participants should guess. Guessing should follow a uniform distribution. Together these processes result in a mixed distribution, consisting of a bell-shaped distribution reflecting the precision (referred to as σ

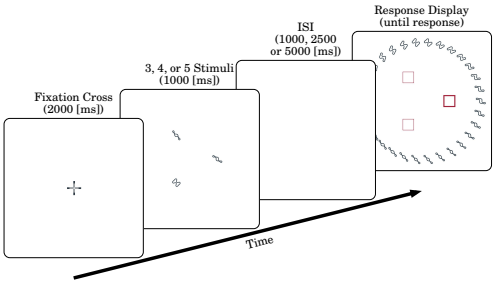

Figure 1: Sequence of a single trial. After the initial presentation of the fixation cross, three, four, or five Fourier descriptors were displayed on a invisible circle for 200 ms. After this a variable interstimulus interval (ISI) of 500, 2500, or 5000 ms followed. At the end of the trial a "shape-wheel" appeared. One of the stimuli positions was highlighted and participants had to indicate the identity of the descriptor on the wheel.

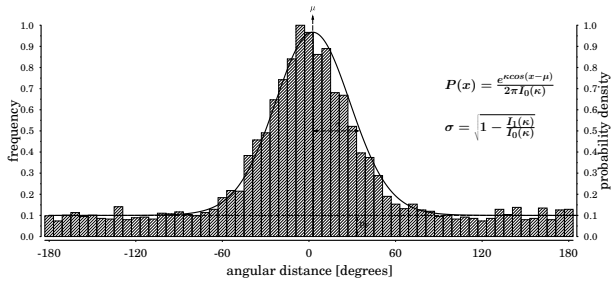

Figure 2: Normalized histogram of exemplary artificially generated data and fitted von Mises distribution.

in the following) and a vertical offset, indicating the probability that participants had no representation of the critical shape (referred to as p_F in the following).

We were interested in the effects of different ISIs on the precision and the amount of stored information, i.e. the variation of σ and p_F. A slot-model would predict a constant σ over time and furthermore that σ would not be affected by the number of stimuli. Once the available slots are filled p_F should strongly increase. A resource model would predict an increasing σ with increasing numbers of items, while p_F should not change.

Participants

16 healthy students (11 males) of the cognitive sciences participated in our experiment, their age ranged between 22 and 33 years (mean age 23.125). All participants reported normal or corrected-to-normal vision and received course credit in exchange for their participation. All participants provided informed consent.

Apparatus

Stimulus presentation and collection of responses were performed by an IBM-compatible computer with a 22-in. display. The stimuli were displayed at a resolution of 1680 by 1050 pixels. The experiment was implemented in C++.

Stimuli

The stimulus set consisted of three, four, or five Fourier descriptors. The phases were chosen pseudo-randomly. The minimal phase difference was 30°. Each descriptor subtended $1.6^\circ \times 1.6^\circ$ degrees of visual angle (viewing distance $\approx$ 70 cm). The descriptors were randomly arranged on an invisible circle with a radius of 4.7° of visual angle. This arrangement was restricted to six possible locations, each spaced by 60°. To obtain the responses a wheel containing 30 shapes evenly distributed over the phase space was presented as well as a cue indicating the critical descriptor (see Fig. 1). The shape wheel was centered on the screen with a radius of 10.2° of visual angle.

Procedure

Each trial started with a fixation cross which lasted for 2000 ms. After this three to five Fourier descriptors were presented for 1000 ms. This was followed by a blank screen which lasted for 1000, 2500, or 5000 ms. Then the response screen was presented and one of the locations was highlighted. The participants responded by clicking on the "shape-wheel". We collected the absolute angle as well as the angular distance to the cued descriptor. Trials lasted until the participants confirmed their response by pressing a key. All 3×3 combinations of number of stimuli and ISI were repeated 30 times, totaling 270 trials. The whole procedure took about 60 minutes.

Results

We applied a quantitative model (Zhang & Luck, 2008) to the data to estimate the probability that a cued descriptor was present in memory ($1-p_F$) as well as the precision (σ) of the representation. We first describe this model, then we report the obtained results.

Data Analysis

According to the model participants have a noisy representation of the crucial descriptor in some trials, in the remaining trials participants are assumed to guess. The noisy representation can be described in terms of a von Mises distribution [1], whereas guessing is modeled as a uniform random process. The resulting mixed distribution is displayed in Fig.2 and can be described as

$$p(x|\mu,\kappa,p_F) = (1-p_F)\frac{e^{\kappa cos(x-\mu)}}{2\pi I_0(\kappa)} + \frac{1}{2\pi}p_F\,, \qquad (1)$$

where μ is the mean of the von Mises distribution, κ denotes the density of the distribution (this can be considered as the inverse of the deviation), I_0 is the modified Bessel function of 0th order, and p_F is the guessing probability.

[1] Due to the circular nature of the response dimension a Gaussian distribution is not feasible.

Later on we will report the deviation of the distribution instead of κ, which can be obtained by

$$\sigma = \sqrt{1 - \frac{I_1(\kappa)}{I_0(\kappa)}}\,, \tag{2}$$

where I_1 is the modified Bessel function of 1st order.

Since the phase of the relevant descriptor varied from trial to trial, the data analysis was based on the angular distance between the response and the phase of the descriptor. Hence the mean of the von Mises distributions should equal 0. As noted above, σ reflects the precision of the representation, larger values of σ indicate a lower quality of the representation. The probability p_F is an indicator for the amount of preserved information. Higher values of p_F indicate less information to be preserved.

The parameters cannot be directly inferred from the observed data. Therefore we fitted the mixed distribution via maximum likelihood estimation.

Estimated Parameters

We concentrated on the estimated values of p_F and σ of the mixed model. We estimated μ, κ, and p_F separately for each participant and each level of the varied factors. The results have to be treated with caution, since the data basis for these fits was quite small, consisting of only 90 samples per fit. For three participants the likelihood values remained comparatively small, indicating that the applied model was not well suited for their data. Hence, the respective data sets were not entered in the analysis.

Fig. 3 shows the obtained results. The estimates with respect to the number of stimuli are plotted on the left panel, whereas the estimates for the different ISIs are plotted on the right panel. Error bars indicate the standard error of the mean.

For a quantitative analysis of the differences of the estimates we performed paired t-Tests. Significant differences on a 5% level are indicated by an asterisk. With respect to the number of stimuli, σ increased significantly. For p_F the estimate was significantly higher in case of four items compared to three items. With respect to the ISI, significant differences were only observed for σ. None of the six μ parameters differed significantly from zero.

The precision of preserved information decreases with the number of items. The amount of preserved information seems less effected, since there was no significant difference between the p_F estimates for three and five items. The precision also seems to change over time, whereas the amount of stored information remains constant. Compared to the effect of the number of stimuli the effect of the ISI is much weaker. The fact that we observed significant changes in the precision of stored information is in conflict with the results reported by Zhang and Luck (2009), but fits the observations of Salmela et al. (2012).

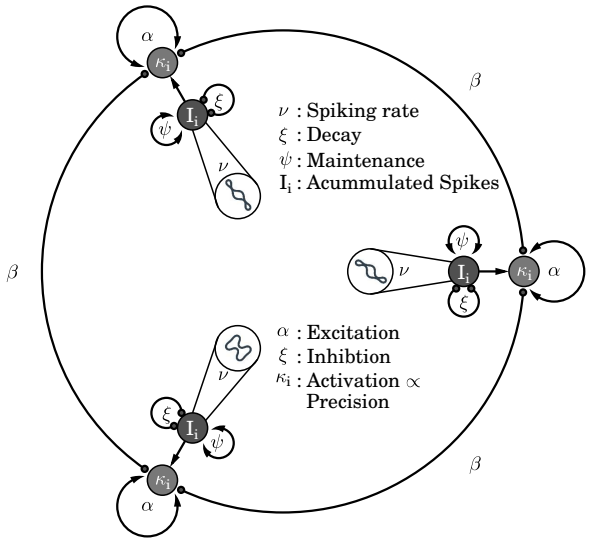

Figure 4: Overview of the model architecture. Input units accumulate spikes from sensory neurons. Memory units receive activation from the input units. Lateral inhibition as well as self-recurrent excitation determine the overall activations of the memory units, which is proportional to the precision of the respective representation. See text for details about the parameters.

Neural Model

So far there are only few models that can account for the temporal dynamics of VWM contents. One of these models is the time-based resource-sharing (TBRS) model (Barrouillet, Bernardin, & Camos, 2004, see Oberauer & Lewandowsky, 2011 for an implementation). TBRS assumes an interplay between temporal decay and a refreshment process to account for dynamic changes in the quality of VWM representations. The encoding stage is neglected however. Another, more neural model is the dynamic field theory (DFT, Johnson, Spencer, & Schöner, 2009), where the dynamic interactions between excitatory and inhibitory layers of neurons are applied to model dynamic changes in VWM content. On the one hand, DFT has a lot of desirable features, for instance capacity limits emerge naturally from the model properties. On the other hand, DFT has a lot of degrees of freedom, rendering direct fitting to observed data rather difficult. Furthermore, the encoding stage is not specified.

We propose a model that can account for encoding of stimuli, as well as for the maintenance of stored representations. We want to achieve this with less degrees of freedom than DFT but still with a neural model. The proposed model is a combination of the theory of visual attention (TVA, Bundesen, 1990) and the short term memory network proposed by Usher and Cohen (1999). Since this network models discrete states, we extended it with the single trace fragility theory (Wickelgren, 1974) to model continuous changes in

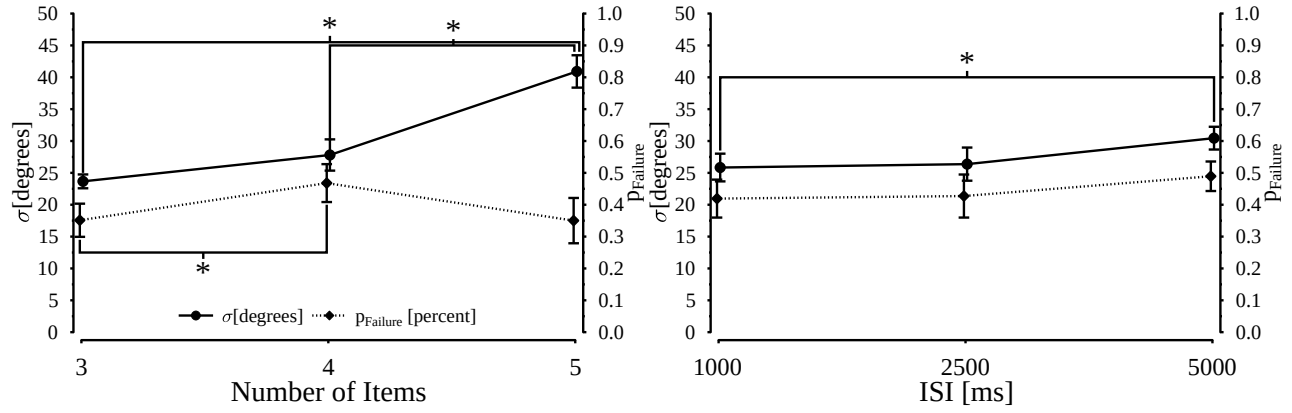

Figure 3: Parameter estimates with respect to the number of stimuli (left panel) and ISI (right panel). Significant differences in the estimates are indicated with an asterisk. Deviation (σ) increases with number of stimuli as well as with ISI. For the failure probability p_F effects were observed for the number of stimuli only.

the precision of stored representations. The input layer of the model accumulates spikes from sensory neurons. The spiking rate of the sensory neurons is based on TVA. Activation from the input layer is forwarded to memory units. The overall activations of the memory units depend on lateral inhibition and self-recurrent excitation. The activity of the memory units varies continuously and is proportional to the precision of the respective representation. Temporal decay is modeled via the leakage of input activation in the absence of sensory input. Initially, the binding between memory units and the input layer is fragile. The longer a representation resides in VWM, the stronger the binding and the weaker the leakage. Fig. 4 gives an overview of the model architecture. The different components of the model are described in the next paragraphs. After this we give a short example of the performance of the model.

Encoding of Stimuli

The encoding stage is modeled via TVA. TVA is a quantitative model of visual encoding that is well suited to account for the selection and categorization of visual stimuli. TVA models visual attention by integrating bottom-up as well as top-down processes.

TVA assumes a competition between different categorizations for incorporation in VWM. This competition can be quantified via a race model, where the rate parameter $\nu(x,i)$ determines the time needed for a categorization of the type "item x belongs to category i" to be finished. The rate parameter depends on the task as well as sensory parameters (see Bundesen, 1990 for details).

We assume a fixed number of sensory neurons that spike during the presentation of stimuli with a rate equal to $\nu(x,i)$ (cf. Bundesen, Habekost, & Kyllingsbæk, 2005). Spikes are accumulated in a separate input layer, which forwards activation to the memory layer that is described in the next paragraph.

Maintenance of Representations

We model VWM in terms of a dynamic neural network (Usher & Cohen, 1999). The activation of each unit is affected by three processes. First, activation is stabilized by self-recurrent excitatory feedback. Second, each unit is inhibited by its neighbors. Third, a unit might receive sensory input that increases its activation (see Fig. 4). We modeled this input in terms of accumulated spikes emitted from the sensory neurons described in the previous paragraph. The dynamics of a single unit in the memory layer can be described via the following differential equation:

$$\frac{d\kappa_i}{dt} = -\kappa_i + \alpha F(\kappa_i) - \beta \sum_{j \neq i}^{N} F(\kappa_j) + I_i + noise\,, \qquad (3)$$

where κ_i indicates the activation of unit i, α is the self-recurrent excitatory weight, β is the weight of the lateral inhibition, I_i denotes the current sensory input supporting unit i, *noise* indicates a uniform noise term, and $F(\kappa)$ is the activation-function (in this case a linear one).

Without sensory input Eq. 3 can be used to model decay of activations to a baseline, given the noise is small enough to prevent random fluctuations. The system has a lot of interesting emergent properties despite its simple structure. It is possible to model serial position effects, capacity limits (with proper choices for α and β), and the development of stable states.

We assume the activation of the memory units to be proportional to the precision of the according representations, which is quantified by the κ parameter of the von Mises distribution in the mixed model. We assume one memory unit per stimulus. If there are more stimuli, lateral inhibition is increased, resulting in an overall reduced activity. The reduced precision observed for higher numbers of stimuli (see Fig. 3, lower panel) emerges naturally. In its original formulation the described network can be used to model discrete states of mem-

ory units, either the activation is above the baseline or not, the transitions are non-linear. Since our data indicates continuous decay of the precision over time, we extended the original model with the single trace fragility theory (Wickelgren, 1974). This extension is described in the next paragraph.

Temporal Decay

As it can be inferred from Eq. 3, the activation of a memory unit depends on the sensory input. Since Buschman et al. (2011) found that neural activation first decays in early areas, we modeled continuous changes in the precision by a decay of the accumulated sensory input [2]. Since older memory traces are more resistant than younger ones (Jost's second law), we assumed this decay to slow down over time.

The single trace fragility theory provides a formal framework for these assumptions, by introducing the concept of *fragility*, which quantifies the susceptibility of stored information for temporal decay. In our model, fragility refers to the binding between input and memory units (see Fig. 4), which modulates the leakage of the input units. Applied to the input term in Eq. 3, decay after display offset can be described by:

$$\frac{dI_i}{dt} = -\xi f I_i \,, \tag{4}$$

where ξ is the decay rate and f denotes the fragility, which is reduced over time:

$$\frac{df}{dt} = -\psi f^2 \,, \tag{5}$$

here ψ denotes the decay rate of the fragility. We assume that fragility starts to decline when a categorization is encoded, hence early categorizations are less prone to temporal decay. Furthermore, we assume that the temporal decay starts after display offset, when sensory information is no longer available to maintain the activity of the memory units. This mechanism concludes the model specification.

Modeling Continuous Representations

To illustrate the functionality of the system, let us assume a simple display containing three stimuli, lets further assume that only one feature dimension of these stimuli is task relevant (e.g. shape) and that the stimuli do not differ with respect to visibility (see Fig. 4, center). In this case the spiking rate of all sensory neurons is equal during the presentation of the stimuli. This process is displayed in the upper panel of Fig. 5. Encoding is a random process, hence the encoding times (t_0) differ. After the offset of the display ($t > D$) the decay mechanism described in Eq. 4 begins to operate. Over time the decay attenuates, since the fragility is reduced (see Eq. 5).

Lateral inhibition between stored representations is visible around time-step 30; the encoding of an additional representation reduces the activation of the previously stored ones.

[2]Please note that Buschman et al. (2011) interpreted this finding in terms of an encoding failure instead of a temporal decay.

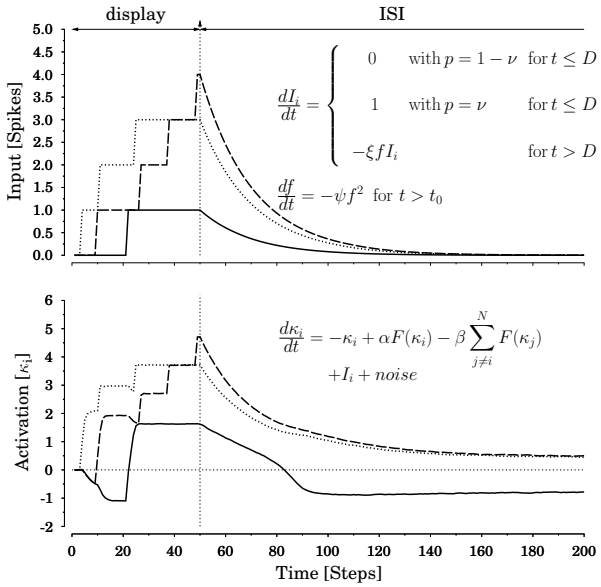

Figure 5: Example of the dynamics produced by the proposed model. Dynamics of the input units are displayed in the upper panel, the overall activation of the memory units is displayed in the lower panel. Parameter D in the upper panel refers to the presentation duration, while t_0 denotes the encoding time of the according representation. Activation of the memory units is assumed to be proportional to the precision of the representation, the decreasing activation reflects the increasing deviation in Fig. 3 (right panel).

After the offset of the stimuli (at time-step 50, vertical dashed line), decay starts. For highly active representations a nearly exponential decay occurs, which attenuates over time, whereas the decay of less active, or older representations can be better described by a power function. As can be concluded from the Fig. 5, it is possible that representations are completely lost (i.e. the according activation falls below zero). The resting level of the system varies, depending on the strength of the lateral inhibition, that is the number of simultaneously active units.

The model so far specifies activations over time; to fit the model to the data in the experiments, we need to convert these activations into probability distributions for memory recollection. We assume the activation of a memory unit to be proportional to the precision of the according representation. Hence, the activations are a measure of the κ parameter of the von Mises distribution. The overall response distribution can be modeled as a mixed distribution, with κ's equal to the activations multiplied with a constant scaling factor (values below zero are considered to be zero). Since a von Mises distribution with $\kappa = 0$ becomes a uniform distribution, p_F can be modeled as well.

Our model is able to account for the encoding of visual stimuli, as well as for changes in the precision of the stored representations. The observed reduction in precision (see

Fig. 3, left panel) due to more stimuli can be accounted for by TVA and lateral inhibition. The change in precision over time is captured by the network dynamics described in Eq. 3 and the temporal decay described in Eq. 4. The apparently high number of parameters is greatly reduced since the TVA parameter ν can be set to one (like in the example). This leaves α, β, ξ, ψ, and the number of sensory neurons per input unit as free parameters. The different components of the model are modular, for instance it is quite simple to model the memory dynamics without Eq. 4. To evaluate the relevance of the different components, we will perform Bayesian model comparisons, once a broader data basis has been acquired.

Conclusions

In this paper we presented data on the temporal changes in the precision of VWM contents. The results indicate that the precision of the representations is strongly affected by the number of stimuli. Precision decreased over time, whereas the amount of preserved information was rather stable (no significant increase of p_F). We proposed a neural network model, which models the encoding as well as the maintenance of information. So far the key findings can be replicated qualitatively, since the overall precision decreases with increasing memory loads as well as over time.

The different approaches to account for VWM properties reflect different paradigms of understanding cognition. While limited-slot models with static precision adhere to the computer metaphor of cognition, resource models are more in line with the dynamic system approach to cognition. The fact that VWM capacity is limited is not questioned, however. Given the findings of Buschman et al. (2011), who observed discrete slots with flexible resources per hemisphere and a possible gating mechanism in the lateral intraparietal cortex, our model offers an implementation of this neural mechanism. As it was noted above, capacity limits emerge naturally in the network due to the interplay of excitation and lateral inhibition, while the individual activations implement a continuous process, modeling the precision of individual representations over time. Thus, the approach integrates slot and resource model perspectives. Further evaluation is necessary to assess the quantitative fit of the model. This requires a broader data basis, involving larger numbers of stimuli to assess the ability of our model to account for higher memory loads, which is usually better accounted for by slot models (Rouder, Morey, Morey, & Cowan, 2011).

References

Anderson, D., Vogel, E. K., & Awh, E. (2011). Precision in visual working memory reaches a stable plateau when individual item limits are exceeded. *The Journal of Neuroscience*, *31*, 1128–1138.

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, *133*, 83–100.

Bays, P. M., & Hussain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*, 851–854.

Berg, R. van den, Shin, H., Chou, W., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *PNAS*, *109*, 8780–8785.

Brady, T., Konkle, T., & Alvarez, G. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision*, *11*, 1–34.

Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, *97*, 523-547.

Bundesen, C., Habekost, T., & Kyllingsbæk, S. (2005). A neural theory of visual attention: Bridging cognition and neurophysiology. *Psychological Review*, *112*, 291–328.

Buschman, T., Siegel, M., Roy, J., & Miller, E. (2011). Neural substrates of cognitive capacity limitations. *PNAS*, *108*, 11252–11255.

Fukuda, K., Awh, E., & Vogel, E. K. (2010). Discrete capacity limits in visual working memory. *Current Opinion in Neurobiology*, *20*, 177–182.

Johnson, J. S., Spencer, J. P., & Schöner, G. (2009). A layered neural architecture for the consolidation, maintenance, and updating of representations in visual working memory. *Brain Research*, *1299*, 17–32.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281.

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.

Oberauer, K., & Lewandowsky, S. (2011). Modeling working memory: a computational implementation of the Time-Based Resource-Sharing theory. *Psychonomic Bulletin & Review*, *18*, 10–45.

Rouder, J. N., Morey, R. D., Morey, C. C., & Cowan, N. (2011). How to measure working memory capacity in the change detection paradigm. *Psychonomic Bulletin & Review*, *18*(2), 324–330.

Salmela, V., Lahde, M., & Saarinen, J. (2012). Visual working memory for amplitude-modulated shapes. *Journal of Vision*, *12*, 1–9.

Usher, M., & Cohen, J. D. (1999). Short term memory and selection processes in a frontal-lobe model. In D. Heinke, G. W. Humphries, & A. Olsen (Eds.), *Connectionist models in cognitive neuroscience* (pp. 78–91). London: Springer-Verlag.

Wickelgren, W. A. (1974). Single-trace fragility theory of memory dynamics. *Memory & Cognition*, *2*, 775–780.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*, 233–235.

Zhang, W., & Luck, S. J. (2009). Sudden death and gradual decay in visual working memory. *Psychological Science*, *20*, 423–428.