Modeling the Temporal Dynamics of Visual Working Memory

J. Lohmann[a,*], O. Herbort[a], M. V. Butz[a]

[a] *Cognitive Modeling, Department of Computer Science & Department of Psychology, Sand 14, Tübingen, 72076 Germany*

Abstract

Visual working memory (VWM) is one of the most crucial parts of the human cognitive system. Research focuses on the apparent limits in the capacity of this system and the reasons for them. So far only a few formal models exist that can account for the temporal dynamics of the amount of information stored in VWM. We propose a combination of the well established theory of visual attention (TVA) with a dynamic memory model, resulting in an iterative, probabilistic framework for VWM. The model includes a consolidation as well as a decay mechanism and employs the strength concept to quantify the availability of a certain memory trace. We evaluate the model on available change detection data.

Keywords: VWM, TVA, Change Detection

1. Introduction

One of the main components of every cognitive task is the storage and maintenance of information in memory. Accordingly, research on working memory has a long tradition in both psychology and neuroscience (for an overview see Wixted, 2004b). Most models of working memory assume distinct systems for the preservation of verbal and visual information (Baddeley, 2003). Change detection tasks are frequently used to study the properties

*Corresponding author. Phone: +49 (0)7071 29-70481

Email addresses: johannes.lohmann@uni-tuebingen.de (J. Lohmann), oliver.herbort@psychologie.uni-wuerzburg.de (O. Herbort), butz@informatik.uni-tuebingen.de (M. V. Butz)

of the visual working memory (VWM). We focus on modeling VWM in this paper.

One way to investigate VWM is to ask participants to detect changes in subsequently presented displays. Early change detection studies (Phillips, 1974; Pashler, 1988) provided evidence that not only the amount but also the persistence of information stored in VWM dynamically changes over time. For example, for short temporal delays between two subsequent displays change detection performance is very accurate, but it deteriorates for longer temporal delays. Pashler also found that change detection performance increased with longer presentation duration of the initial display. Apparently this is due to the fact that more information can be encoded the longer the initial display lasts. On the other hand it is possible that the encoded information becomes more stable for longer display durations. Seeing that VWM and visual perception additionally appear to be highly intertwined (Alvarez and Cavanagh, 2004; Gao et al., 2011), we approximate the dynamics of VWM in conjunction with visual perception by modeling encoding and memory consolidation with the same hypothetical process.

By now, only a few quantitative models are available that describe the process of stimulus encoding and the preservation of the obtained information at once (a neural model was proposed by Johnson et al., 2009). Here, we provide a parsimonious quantitative model that can account for changes in the amount of stored information over time. To test the idea that memory encoding and VWM maintenance processes interact, we introduce a memory mechanism that also operates during the display presentation.

In the next section we give an outline of the model and sum up the underlying assumptions. Then we give a short description of TVA. After this we describe the features of our memory model in detail and give an example for the predictions of the complete model. We use the results of Phillips (1974) to evaluate our model. A short discussion concludes the paper.

2. General Assumptions

We investigate processing of visual stimuli at the stage of "perceptual units" (Bundesen, 1990). Perceptual units can be considered as segmented parts of the current visual input. Each unit can be described by feature dimensions like color or shape. The encoded categorizations of the feature dimensions are assumed to be the information that is stored in VWM. The theory of visual attention (TVA), proposed by Bundesen (1990), captures this encoding stage in a formal framework.

We assume that every encoded categorization of a certain stimulus dimension can be described with a *strength* that quantifies the availability of the respective categorization. This strength is not constant but changes over time. Two processes affect the strength. First a consolidation process increases the strength over time. When the strength of a certain categorization increases during the presentation of a display, we refer to this process as on-line consolidation. Otherwise, we refer to consolidation as off-line consolidation. Seeing that consolidation generally takes place all the time, categorizations that are encoded earlier typically reach higher strength values than comparable types of categorizations that are encoded later on. Additionally, consolidation depends on memory load. Second a degradation process reduces the strength over time. We assume the degradation to take place at a constant rate after the offset of a certain stimulus display. Moreover, we assume this process to be independent of memory load. If the strength of a certain categorization falls below a threshold the respective categorization is removed from VWM. This is similar to the idea of *some-or-none* representations proposed by Zhang and Luck (2009).

To sum up our model describes the following processes:

a) Encoding of categorizations of stimulus dimensions (e.g. color or shape) from a display.
b) On-Line consolidation of the stored categorizations during the presentation of the display.
c) Off-Line consolidation of the stored categorizations after the offset of the display.
d) Decay of stored categorizations after the offset of the display.

To test the resulting model the estimated memory load is transferred into a behavioral measure. We focus on modeling change detection performance. In the next sections we describe the different assumptions in more detail and provide the necessary formalizations.

3. Encoding of Information

Before we can consider the properties of information stored in VWM it is necessary to describe how information is encoded in the first place. TVA is a quantitative model of visual attention that accounts for a broad range of phenomena (Bundesen et al., 2005; Logan, 2002). TVA was successfully applied to model iconic memory (Sperling, 1960), visual search (Treisman and Gelade, 1980), switch costs (Logan et al., 2007), as well as attention deficits in clinical populations (Duncan et al., 1999). TVA allows quantitative predictions of the amount of information stored in VWM at a certain time. To model visual attention TVA integrates bottom-up processes (via the sensory properties of the relevant information) and top-down processes (via the intention to perform a task).

TVA models the encoding of visual stimuli in VWM as the combination of a *filtering* and a *pigeonholing* process. Filtering selects objects, whereas pigeonholing assigns categories to the selected objects. TVA proposes a race model where different categorizations compete for the incorporation in VWM. This race is formalized as the conditional probability of a categorization to be encoded, given that is was not encoded earlier. The following *rate equation* describes the probability that the categorization of item x belonging to category i enters VWM:

$$\nu(x,i) = \eta(x,i)\beta_i \frac{w_x}{\sum_{z \in T} w_z} \,, \tag{1}$$

where the categorization likelihood $\nu(x,i)$ depends on the sensory evidence $\eta(x,i)$ that object x belongs to category i, on the perceptual decision bias β_i for category i, and on the attentional weight w_x relative to all other attentional weight values w_z for all objects z in the visual display T.

The attentional weights w_z depend on the *pertinence* of a given categorization that can be considered as the subjective relevance of this categorization. For every object x in the visual field, an attentional weight is obtained by:

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

where R denotes the set of all perceptual categories, $\eta(x,j)$ denotes the sensory evidence for element x belonging to category j, and π_j is a pertinence value for category j. The higher the pertinence of a certain category, the higher the likelihood to attend to objects that fall into the respective category.

TVA realizes filtering by attentional weights. If the task is to select red objects the value of π_{red} and the resulting weights for red objects would be high. Pigeonholing is biased by parameter β. If a task requires the categorization of letters rather than digits, the value of β would be higher for letters. The consequence of combined filtering and pigeonholing in this example would be the faster encoding of red letters compared to other stimuli.

Despite examples of successful applications, TVA cannot account for decay of information stored in VWM. Bundesen assumed a fixed capacity VWM model that is filled by a Bernoulli process. As the capacity is fixed, it cannot account for any loss of stored information over time. Therefore we propose a dynamic model of VWM, which also considers memory decay.

4. A dynamic Model of VWM

We assume a continuous mnemonic resource (Bays and Hussain, 2008; Wilken and Ma, 2004; Verghese, 2001) that is used to consolidate information in VWM[1]. Memory limits emerge over time as it becomes more and more difficult to maintain the stored information. To describe the state of certain stored categorizations, we employ the *strength* concept proposed by Wickelgren (Wickelgren, 1970, 1974). According to this theory the availability of a certain memory trace can be described by its strength. This strength changes over time. Initially memory traces are weak but their strength increases over time. Hence older traces are more stable than younger ones (Jost's second law, see Wixted, 2004a). For short retention intervals even the weak traces can be preserved. For longer intervals only the strongest traces persist.[2]

Initially the strength of an encoded categorization is set to 1.0. We assume three different processes that affect the strength of a certain stored categorization over time. First, a decay process applies to each stored categorization after stimulus display offset. Second, an on-line consolidation mechanism increases the strength of a categorization after its encoding throughout the presentation of a display. This assumption is based on Jost's second law to give older traces the proposed advantage in durability. Third, off-line consolidation increases the strength after display offset.

Some of these assumptions are shared with another recent model of VWM. The *time-based resource-sharing model* (TBRS), proposed by Barrouillet

[1]Please note that the idea of fixed slots (Zhang and Luck, 2008) is a prominent alternative to this approach (for a comparison of both approaches see Fukuda et al., 2010).

[2]Please note that Wickelgren assumed an ever decaying strength, an increase of strength that is realized by the consolidation components of our model was never proposed by Wickelgren. Instead he assumed the decay to slow down over time.

et al. (2004) also assumes a close interaction between processing and maintenance of information. Furthermore, Barrouillet et al. assume that unattended traces suffer from temporal decay. Compared to our model TBRS does not account for the encoding of information and provides no formal approach to the mentioned maintenance and decay mechanisms (a recent implementation of TBRS can be found in Oberauer and Lewandowsky, 2011).

So far we have specified a model that assumes a close relation between encoding and maintenance of information in VWM. We assume temporal decay to be the primary cause of degradation of information. Compared to other models of VWM we assume a continuous mnemonic resource instead of a fixed number of slots. As it was noted earlier it is still in discussion if VWM is better described in terms of a limited resource or a fixed number of slots. But see van den Berg et al. (2012) for a recent comparison of resource-based and fixed capacity models of VWM.

We are now turning to the formal description of the mentioned processes.

4.1. Decay

We assume the strength of every stored categorization to degrade at a constant pace after the offset of the stimulus display. The decrement is denoted as ξ, it is assumed to be a random variable and to be unaffected by the load of VWM. If a certain strength falls below 0, the according categorization is lost.

4.2. On-Line Consolidation

We assume the conditional encoding probability $\nu(x, i)$ to be proportional to the amount of mnemonic resources that are used to consolidate a certain categorization after it enters VWM. This assumption is justified as $\nu(x, i)$ reflects top-down influences, such as the relevance of a certain categorization (see Eq. 1). Accordingly, we assume the strength $s(x, i)$ to increase by $\nu(x, i)$, each time consolidation applies. We furthermore assume that only one categorization is consolidated at a time, essentially assuming a serial consolidation process similar to the one discussed in Schneider (1999). The categorization that is consolidated next is chosen randomly (with replacement). Each consolidation may require several iterations.

4.3. Off-Line Consolidation

A similar mechanism like the proposed on-line consolidation is assumed to operate after the offset of the display. The difference between on-line and off-line consolidation is the value of the increment of a certain strength $s(x, i)$. We assume that the sum of all conditional encoding probabilities to be proportional to the amount of mnemonic resources, referred to as κ, that can be used for consolidation:

$$\kappa \propto \sum_{x \in T} \sum_{i \in R} \nu(x, i) \tag{3}$$

For off-line consolidation, we assume that this amount is distributed over the currently stored categorizations. If all categorizations are of equal relevance, the amount of κ used to consolidate a single categorization can be obtained by dividing κ by the number of currently stored categorizations. In effect, the consolidation of a single categorization is more effective when less categorizations are stored in VWM. As a consequence, the assumed decay mechanism bears a stronger influence if VWM load is high. In the next section we give an example for the predictions of our integrated model.

5. Example

As an example, lets assume a very simple stimulus display containing three objects with different colors. Lets further assume that the colors of these objects are equally relevant for the current task, whereas other possible features are irrelevant. We now want to use our model to predict the temporal changes in VWM content. An exemplary complete time-course is displayed in Fig. 1.

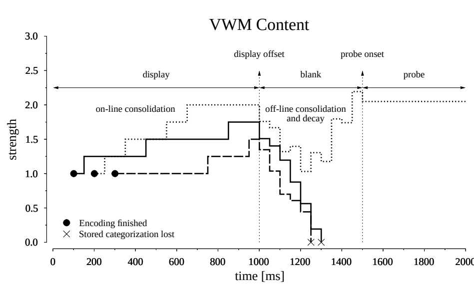

Figure 1: Example of the predicted changes in VWM over time.

5.1. Encoding

As only color information is relevant all the respective $\beta(x, color)$ and $\pi(x, color)$ would be high, whereas all other β and π values would be 0. At every time step each of the color categorizations is encoded with a certain probability ($\nu(x, color)$) and stored in VWM. As this is a probabilistic process the time of successful encoding differs for the different categorizations (see Fig. 1, the time of encoding is indicated by a filled circle).

5.2. On-Line Consolidation

As it is displayed in Fig. 1 the strength of the stored categorizations increases over time during the presentation of the display. If only one categorization is stored it is consolidated every time consolidation takes place. As only one categorization is consolidated at a time, the growth of an individual strength value declines as more and more categorizations are encoded.

5.3. Decay

After the offset of the display the strength of the stored categorizations decays. As the decay is noisy it differs from time-step to time-step. As it is displayed in Fig. 1, two categorizations are lost after about 300 ms due to the decay process.

5.4. Off-Line Consolidation

After the offset of the display the decay in strength is encountered by an consolidation mechanism. It is impaired for higher memory loads. As more and more categorizations are lost it becomes more effective and can protect the remaining categorizations from further decay (see Fig. 1, at around 1300 ms). We did not implement an upper bound for the strength hence it can increase infinitely.

6. Results

In this section we apply our model to the data from the first experiment reported by Phillips (1974). This study investigated change detection performance for lighted pixels in a square matrix. Each pixel in the matrix had a chance of 50% to be lit. The matrix was presented for one second. The matrix size varied between 4×4, 6×6 and 8×8 pixels [3]. After an inter stimulus interval (ISI) of either 20, 1000, 3000 or 9000 ms, a probe display appeared. In 50% of the trials the probe display was equal to the initial display, in the other trials it differed with respect to one pixel. The participants had to indicate if the probe display differed from the initial one. Phillips (1974) collected the percentage of correct responses as a dependent measure. The results suggested two different types of storage systems. First, a high-capacity but short lasting iconic storage system. Second, a more persistent but capacity limited short term storage system. The data basis for the evaluation is quite small comprising only 12 mean detection probabilities. Nevertheless the data pattern is challenging because of the broad variation in the length of ISIs, covering sensory memory as well as short term memory. This is the main reason for which we decide to use this data set instead of more contemporary studies like the one reported by Vogel et al. (2001).

To obtain predictions of the percentage of correct responses we have to transfer the estimated VWM load after the ISI into a probability. The predicted probability of success would equal the number of available categorizations for the relevant feature dimension:

$$p_{detect} = \frac{E_{preserved}}{T} \tag{4}$$

Where T refers to the number of relevant categorizations and $E_{preserved}$ is the number of stored categorizations in the relevant dimension. As the participants had to perform a same / different judgment, a correction for guessing is applied. Usually the guessing probability should be 0.5:

$$p_{predicted} = p_{detect} + \frac{(1 - p_{detect})}{2} \tag{5}$$

This is similar to the formula proposed by Pashler (1988), except for the fact that we assume a constant guessing probability for all participants.

We used four different versions of our model to investigate if the different mechanisms improved the predictions. The first version modeled encoding

[3] Please note that this stimulus material is problematic as the participants might be able to store geometric patterns instead of individual pixels. Hence the storage capacity of VWM should not be inferred from the reported detection probabilities. The temporal change in available information is valid however and was reproduced for instance by Pashler (1988).

with TVA and assumed constant decay during the ISI. Neither on-line consolidation nor off-line consolidation were used. The second version included the off-line consolidation process during the ISI. The third version included the on-line consolidation mechanism. The fourth version applied both off-line and on-line consolidation. All η, β and π values were set to 1.0. Hence the ν values only differed between matrix sizes. The equality of the TVA parameters is plausible because the categorization of every pixel as either lit or not lit was equally relevant for the task. Due to our assumption of proportionality between κ and ν, only ξ was treated as a free parameter. We applied equally distributed noise to ξ, accordingly the applied decay rate could vary between 0 and ξ in each time step. As noted earlier we assume consolidation to take time. Hence all models that applied on-line, or off-line consolidation or both had an additional free parameter, referred to as *lag* that determined how much iterations have to pass between two consolidation phases. All free parameters were constant over matrix sizes. As encoding, consolidation and decay are modeled as probabilistic processes, we averaged 50 independent runs for every iteration of the parameter estimation.

The predictions of the different versions are displayed in Fig. 2. Parameter estimation was done with a downhill simplex simulated annealing algorithm, minimizing the root mean square error (RMSE) of the predictions. Due to the probabilistic nature of our model, 50 independent runs were aggregated in every optimization step. The results presented in Fig. 2 and Table 1 are the average of 50 independent runs, applying the best parameter sets for the different setups with respect to RMSE.

Table 1: Parameter estimates, RMSEs and r^2 values for the different setups.

Setup	ξ	lag	RMSE	r^2
Decay only	0.020	-	0.154	.54
Off-Line consolidation	0.031	5	0.048	.92
On-Line consolidation	0.021	0	0.151	.58
Full model	0.030	5	0.046	.93

Results with respect to RMSE and declared variance (r^2) are displayed in Table 1. Additionally the parameter estimates are included. The best results were obtained with models assuming decay and off-line consolidation. The

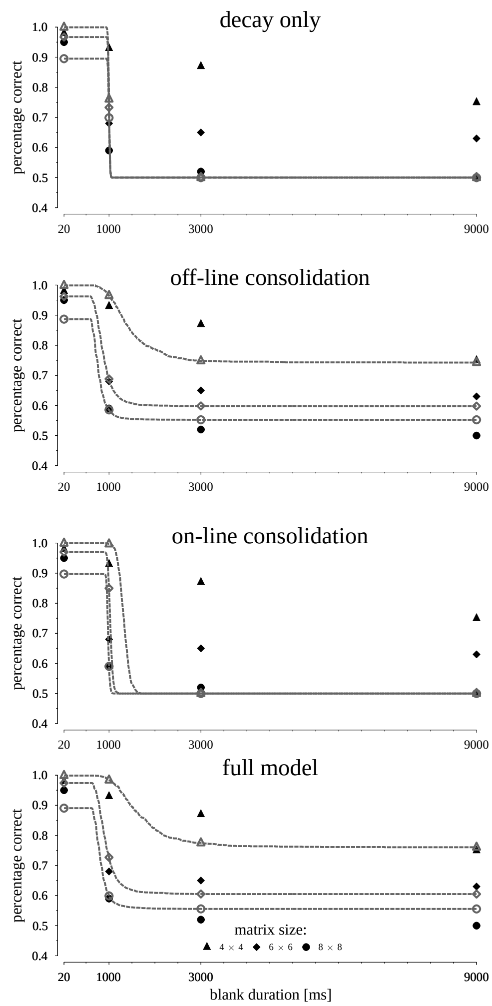

Figure 2: Results obtained with the different versions of our model. Different markers indicate the different matrix sizes. The predicted probabilities are indicated by dashed lines, the predictions for the measured ISIs are indicated by empty markers.

addition of the on-line consolidation improved the fit only slightly. At least for this data pattern it seems not necessary to assume an on-line consolidation mechanism. This can also be concluded from the parameter estimates. There are nearly no differences between the estimate of ξ between the model that only applied off-line consolidation and the full model, applying both on-line and off-line consolidation.

7. Discussion

We proposed a model that accounts for the temporal dynamics of the information stored in VWM. Therefore we combined TVA with a dynamic memory model that assumes the concurrent operation of a consolidation and a decay mechanism. Our model combines the encoding mechanism proposed by TVA with the strength concept developed by Wickelgren. The assumed on-line consolidation was included to account for the age of different memory traces and is in line with Jost's second law. The modeling of the first experiment reported by Phillips (1974) indicates that this mechanism is not necessary to account for the results. Possibly it becomes more relevant for longer display durations. This would be plausible since Jost's law was based on observations of long term memory recall. However some aspects of the model need further clarification.

First our model is neutral about the source of degradation. The original strength theory proposed by Wickelgren assumed interference between successive stimuli to be the main source of memory degradation. The decay mechanism assumed in our model is more in line with trace decay due to prolonged retention intervals. To account for interference it would be necessary to account for the onset of successive stimuli. As it is displayed in Fig. 1, currently the presentation of subsequent displays is not supposed to affect the VWM content. It remains open if the current architecture is able to account for findings that highlight the role of interference for VWM contents (Makovski et al., 2008).

Second the proposed on-line consolidation mechanism was not necessary for modeling the data reported by Phillips (1974) accurately. Possibly the display duration of 1000 ms applied by Phillips is too short to require the assumption of on-line consolidation. Therefore the model should be applied to change detection data that was obtained with longer display durations to further investigate the validity of the proposed on-line consolidation mechanism. Possible data for evaluation could be obtained from the experiments reported by Hollingworth and Henderson (2002).

Third the data mentioned so far is quite old and the pattern is rather simple. Evaluation of the model based on data from the literature can only provide limited evidence for the validity of the underlying assumptions. Especially the lack of information about distributions of detection probabilities prevents a detailed maximum likelihood analysis, which would allow comparisons between different setups of the model. Therefore we are currently conducting change detection experiments, manipulating exposure duration as well as ISI's based on predictions from our model to investigate if the predicted detection probabilities are reproduced by the participants. Additionally we are collecting the scan paths of the participants with an eye-tracker to obtain a more accurate measure of the actual memory content as well as an estimate of the time used for refreshing existing representations. With these data we will be able to provide a deeper evaluation of our model.

Even with the mentioned shortcomings the results obtained with the model are promising. As the amount of free parameters is very small it is highly unlikely that this is due to the flexibility of the model. But especially the necessity of the assumed on-line consolidation mechanism remains unclear and has to be investigated in more detail in the future. Further extensions of the model could include a more detailed account for early visual processing, for instance by a layered battery of Gabor filters as it was proposed by Serre et al. (2007). Furthermore the mechanisms by which the top-down control variables (β and π) are assigned could be modeled in more detail. One possible approach would be a Bayesian model like the one described by Chikkerur et al. (2010). Finally, the memory model could be reformulated in terms of a neural network. One possible architecture would be the one proposed by Usher and Cohen (1999). This neural network is able to account for changes in the strength of representations over time and can account for recency effects as well as effects of presentation rate. Compared to other architectures the number of free parameters is small, as only an estimate for the self-recurrent excitation and the strength of the lateral inhibition is required. Both parameters can be related to the parameters of our model.

Hopefully a further integration of neural models will provide another bridge between cognitive psychology and neuroscience, combining the desirable features of both approaches: A detailed description of the involved processes and an output format that allows direct evaluation with observed data.

References

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.

Baddeley, A., 2003. Working memory: Looking back and looking forward. Nature Reviews Neuroscience 4, 829–839.

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.

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.

Chikkerur, S., Serre, T., Tan, C., Poggio, T., 2010. What and where: A bayesian inference theory of attention. Vision Research 55, 2233–2247.

Duncan, J., Bundesen, C., Olson, A., Humphreys, G., Chavda, S., Shibuya, H., 1999. Systematic analysis of deficits in visual attention. Journal of Experimental Psychology: General 128, 450–478.

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

Gao, T., Gao, Z., Li, J., Sun, Z., Shen, M., 2011. The perceptual root of object-based storage: An interactive model of perception and visual working memory. Journal of Experimental Psychology: Human Perception and Performance 37, 1803–1823.

Hollingworth, A., Henderson, J. M., 2002. Accurate visual memory for previously attended objects in natural scenes. Journal of Experimental Psychology: Human Perception and Performance 28, 113–136.

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.

Logan, G. D., 2002. An instance theory of attention and memory. Psychological Review 109, 376–400.

Logan, G. D., Schneider, D. W., Bundesen, C., 2007. Still clever after all these years: Searching for the homunculus in explicitly cued task switching. Journal of Experimental Psychology: Human Perception and Performance 33, 978–994.

Makovski, T., Sussman, R., Jiang, Y. V., 2008. Orienting attention in visual working memory reduces interference from memory probes. Journal of Experimental Psychology: Learning, Memory, and Cognition 34, 369–380.

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

Pashler, H., 1988. Familiarity and visual change detection. Perception & Psychophysics 44, 369–378.

Phillips, W. A., 1974. On the distinction between sensory storage and short-term visual memory. Perception & Psychophysics 16, 283–290.

Schneider, W. X., 1999. Visual-spatial working memory, attention, and scene representation: A neuro-cognitive theory. Psychological Research 62, 220–236.

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., Poggio, T., 2007. Robust object recognition with cortex-like mechanisms. IEEE Transactions on Pattern Analysis and Machine Intelligence 29, 411–426.

Sperling, G., 1960. The information available in brief visual presentations. Psychological Monographs 74(I 1, Whole No. 498), 1–29.

Treisman, A., Gelade, G., 1980. A feature-integration theory of attention. Cognitive Psychology 12, 91–136.

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

van den Berg, R., 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.

Verghese, P., 2001. Visual search and attention: A signal detection theory approach. Neuron 31, 523–535.

Vogel, E. K., Woodman, G. E., Luck, S. J., 2001. Storage of features, conjunctions, and objects in visual working memory. Journal of Experimental Psychology: Human Perception and Performance 27, 92–114.

Wickelgren, W. A., 1970. Time, interference, and rate of presentation in short-term recognition memory for items. Journal of Mathematical Psychology 7, 219–235.

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

Wilken, P., Ma, W. J., 2004. A detection theory account of change detection. Journal of Vision 4, 1120–1135.

Wixted, J. T., 2004a. On common ground: Jost's (1897) law of forgetting and Ribot's (1881) law of retrograde amnesia. Psychological Review 111, 864–879.

Wixted, J. T., 2004b. The psychology and neuroscience of forgetting. Annual Review of Psychology 55, 235–269.

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.