Visual working memory for natural scenes: the effect of chromatic, luminance and spatial frequency content

Ben J. Jennings[1,3*], Joy T. W. Tseng[2] and Frederick A. A. Kingdom[3]

[1]Centre for Clinical and Cognitive Neuroscience, Division of Psychology, College of Health Medicine and Life Science, Brunel University London, U.K.

[2]Department of Microbiology & Immunology, McGill University, Montreal, Canada

[3]McGill Vision Research, Department of Ophthalmology, Montreal General Hospital, Montreal, Canada

*Corresponding author (ben.jennings@brunel.ac.uk)

Abstract

Long-term memory for images of natural scenes is known to be very good. However visual working memory (VWM) for natural scene stimuli is less well understood. We investigated VWM for natural scene stimuli by measuring VWM performance as a function of both encoding time and cognitive load level, employing a method that approximates everyday natural vision. VWM performance was compared between (a) scenes containing either full chromatic and luminance information, (b) luminance-only (isochromatic) information and (c) chromatic-only (isoluminant) information. VWM performance was also measured for scenes in which the scene's structure had been destroyed by Fourier phase scrambling, or following removal of either the high or low spatial frequencies. It was found that recall ability for isoluminance scenes was relatively poor, as it was also for the phase scrambled scenes with high cognitive load or short encoding time. However, recall ability was similar for the full colour (i.e., chromatic and luminance information combined) and luminance-only scenes, except for very brief presentation times where performance for the luminance-only scenes was worse. These findings suggest that spatial scene structure is important for good VWM performance, and for very brief presentations there is a particular reliance on chromatic information.

Introduction

Humans have a remarkable ability to remember images of scenes over relatively long periods of time (Isola et al. 2014). This long-term-memory ability, however, begs the question: what of working memory for natural scene images. Working memory refers to the cognitive mechanism that is capable, within limits, of temporarily storing information in such a manner as for it to be readily available for retrieval and manipulation. This ability can hence assist, for example, in making decisions that may affect a current task (Miyake and Shah, 1999). The current study investigates visual working memory (VWM), a vital function that facilitates the transient encoding of incoming visual information for rapid decision making. For example, VWM is useful for keeping track of previously inspected regions of an unfamiliar terrain while navigating within it. VWM however suffers from a limitation in holding large quantities of information. Miller (1956) famously demonstrated the "magical number" of 7±2 as a typical number of items that can be retained in working memory at a given time. Other studies have reported slightly different values and shown significant individual performance differences, but in general the maximum number of items is relatively low (for example; Cowan 2000; Daneman and Carpenter 1980).

In the current study we investigate VWM performance for natural scene images: 'raw' scenes, i.e., containing both colour (chromatic) and luminance information, referred to as "colour" scenes, luminance-only scenes, i.e. grayscale images containing no colour information and referred to as "luminance isolating" scenes, and colour-only scenes, i.e. scenes containing no luminance information and referred to as "isoluminant" scenes.

Previous colour memory studies investigating VWM have typically employed artificial stimuli. For example, in an electrophysiological study by Kosilo, Haenschel, and Martinoivc (2015), memory performance for isoluminant and luminance-only arbitrarily-curved shape stimuli was compared, using a delayed match-to-sample task. Kosilo *et al.* showed that the amplitude of the early visual P1 component was highly correlated with memory performance. Performance was worse at isoluminance compared to the luminance condition and performance rapidly deteriorated under high-load conditions.

Colour memory studies of natural scenes have focused on longer-term memory not VWM, often due to the nature of the employed task, e.g., presenting images in an encoding phase followed by a query phase. Wichmann, Sharpe, and Gegenfurtner (2002) compared memory for colour and luminance-only images of natural scenes. In the encoding phase of the experiment they presented 48 images, with a 7 s blank interval between images. This was followed by a query phase where observers classified the same 48 test images plus an additional 48 distractor images (presented in a random order) as either having been seen or not. Wichmann *et al.* showed a ~5-10% superiority in the recall of the colour compared to luminance isolating scenes, independent of the initial scene exposure duration (50 to 1067 ms). They also showed that recall performance for both colour and luminance scenes plateaued at, or above, 40% contrast. This contrast is lower than that normally experienced with real scenes, so it was concluded that the performance colour versus luminance performance difference was not due to any differences in chromatic and luminance scene contrast. Wichmann et al. also showed that performance deteriorated if the scenes were initially in full colour and subsequently tested as luminance-only, or vice versa, suggesting that the chromatic content of scenes is part of the memory representation. They also ruled out attentional factors as the cause of the performance differences, a claim later corroborated by (Marx, Hansen-goos, Thrun and Einhauser, 2014). Finally, Wichmann *et al*. showed that the colour advantage could be destroyed if the scenes contained false colour information, suggesting that object colour familiarity was important for scene memory (e.g. see Oliva and Schyns (2000)). The superiority of colour compared to luminance-only images for long-term recognition memory has been confirmed by Spence et al. (2006). These studies inevitably raise the question as to whether similar results pertain for VWM.

Besides colour and luminance, spatial frequency is a dimension of interest in memory studies of natural scenes. In general, high frequencies in an image capture the featural details such as edges, while the low frequency content contains both configuration information, i.e. how those features are arranged, as well as surface information, e.g. colour, brightness texture, etc. (Wenger and Townsend 2000). Magnussen and Dyrnes (1994) compared discrimination thresholds for gratings as a function of the inter-stimulus interval (ISI), and found perfect recall for ISIs ranging from 1 second to 50 hour, and for each frequency tested (2.5, 5 and 10 cpd). They suggested that spatial frequency is encoded by a mechanism specialised for pre-categorical storage of visual features. These experiments however only employed simple sine-wave gratings; in the current study we use images of natural scenes.

To summarise: we have compared VWM performance for colour, luminance and isoluminant images of natural scenes, for different exposure times and with either the high or low spatial frequency ranges removed (see Fig. 1 for examples of stimuli). We also compared VWM performance for scenes with and without structure, the latter achieved by Fourier phase scrambling.

General Methods

Observers

In total 46 naive observers took part in the experiments, aged 23±5 (mean ± SD) years. All had normal colour vision and possessed 6/6 vision, some with optical correction. Each observer provided written consent before testing commenced. All experiments were approved by the McGill Ethics Committee and were performed in accordance with the declaration of Helsinki.

Equipment

Stimuli were presented on a CRT Sony Multiscan Trinitron G400 monitor, controlled by a ViSaGe system (CRS ltd, UK), and controlled by a Dell Precision T1650 host computer. All experimental software was custom written using MatLab (MathWorks Inc). The display was gamma corrected using a colorCAL 123 (CRS Ltd., UK) controlled via the vsgDesktop software. The CIE (x, y, Y) coordinates of the red (R), green (G) and blue (B) phosphors, as measured using a SpectroCAL spectroradiometer (CRS Ltd., UK), at maximum luminance outputs, were R (0.62, 0.34, 16.6 cd m^{-2}), G (0.28 ,0.61, 55.4 cd m^{-2}), and B (0.15, 0.07, 9.6 cd m^{-2}). The stimuli were presented on a mean gray background with RGB: 0.29, 0.31, 40.6 cd m^{-2} corresponding to (0.5, 0.5, 0.5) in RGB colour space. The monitor was run with a refresh rate of 100 Hz and a resolution of 1280 × 960 pixels, with one pixel measuring ∼0.94 × 0.94 minute of arc at the viewing distance of 100 cm.

Stimuli

All stimuli consisted of images of natural scenes. The raw digital photographs came from the McGill Calibrated Color Image Database (Olmos and Kingdom 2004), as well as images taken by the database cameras but not yet uploaded onto the database. This resulted in 1260 calibrated images.

Image processing was performed using MatLab (Mathworks, 137 Natick, MA). Images were gamma corrected, as described in Olmos & Kingdom (2004). The actual stimuli were pseudo-randomly selected square subsections of each image, with each stimulus subtending 512 x 512 pixels. The luminance and isoluminant images were generated by converting the gamma corrected images from RGB space to the Y'UV colour space, using a 3 x 3 RGB to Y'UV transformation matrix. The Y'UV space contains three layers: Y' contains the luminance information while the U and V layers contain the colour information. To generate the luminance images the U and V layers were set to zero, removing the chromatic information. To generate the isoluminant images each pixel in the Y' layer was set to the mean value of that layer, in a method analogous to (Harding and Bloj 2017). The isoluminant or luminance Y'UV images were then transformed back into RGB space using an inverse 3 x 3 matrix (Y'UV to RGB). These square stimuli were finally made round by applying a thin circular Gaussian edge, this created a "soft" boundary against the mean-grey background.

Examples of the colour, luminance and isoluminant stimuli are shown in the top row of Fig. 1, while the second row shows examples of the same conditions after Fourier phase scrambling. The phase scrambled stimuli were generated using the method outlined in Yoonessi, Kingdom, and Alqawlaq (2008). With this method, the absolute phases of each of RGB layer were scrambled, but their relative phases were preserved, ensuring that the chromatic content of the image was preserved while its structure was destroyed. The method employed a 2D (two-dimensional) fast Fourier transform implemented in MatLab to extract the amplitude (*A*) and phase spectra (*P*), as expressed in Eq. 1 and 2, respectively. Here, F_r and F_i represent the real and imaginary components of the Fourier transform as a function of frequency (ω) in the y and x directions.

$$A = \sqrt{F_r(\omega_x, \omega_y)^2 + F_i(\omega_x, \omega_y)^2} \qquad \text{Eq. 1}$$

$$P = \arctan\left(\frac{F_i(\omega_x, \omega_y)}{F_r(\omega_x, \omega_y)}\right) \qquad \text{Eq. 2}$$

To produce stimuli that isolated the low and high frequency information of the scenes a wavelet based procedure was performed. Each image was filtered using a bank of Log-Gabor filters at four orientations. Examples of these low and high frequency images are shown in the bottom two rows of Fig. 1, for the colour, luminance and isoluminant conditions.

Testing procedure

Visual working memory was measured using an n-back paradigm (Kirchner, 1958). This roughly mimics the everyday visual experience of traversing the world and is hence ecologically different from methods in which a set of stimuli are first presented serially (an encoding phase) and then later observers are tasked with identifying previously seen images amongst new distractor images (decoding phase).

Each testing block comprised an image stream of 96 images. A trial was two seconds in duration, with images presented at the start for either 1000 or 30 ms. with the remainder of the trial (either 1000 or 1970 ms, respectively) a blank uniform mid-grey. In each block 12 target images were presented at random positions in the image stream. Each target image was presented again either after 0, 1 or 2 distractor images between their first and second presentations or (with a different set of observers) 3, 4, or 5 distractors images between them, more distractors corresponding to a higher cognitive load. Across all blocks observers were never presented with the same distractor image more than once, ruling out the possibility of remembering scenes from previous blocks. During the phase scrambled blocks the target images employed exactly the same randomization for both the first and subsequent presentation.

Observers were instructed to respond on every trial, at any time during the trial, with one of two possible responses indicating either (i) the current image had not been seen before, or (ii) it had been seen before. If a response was not given before the subsequent trials onset it was recorded as an incorrect response. Fig. 2 shows an example subsection of a real colour image stream, with a scene of purple flowers as the target, presented with distractor level 2.

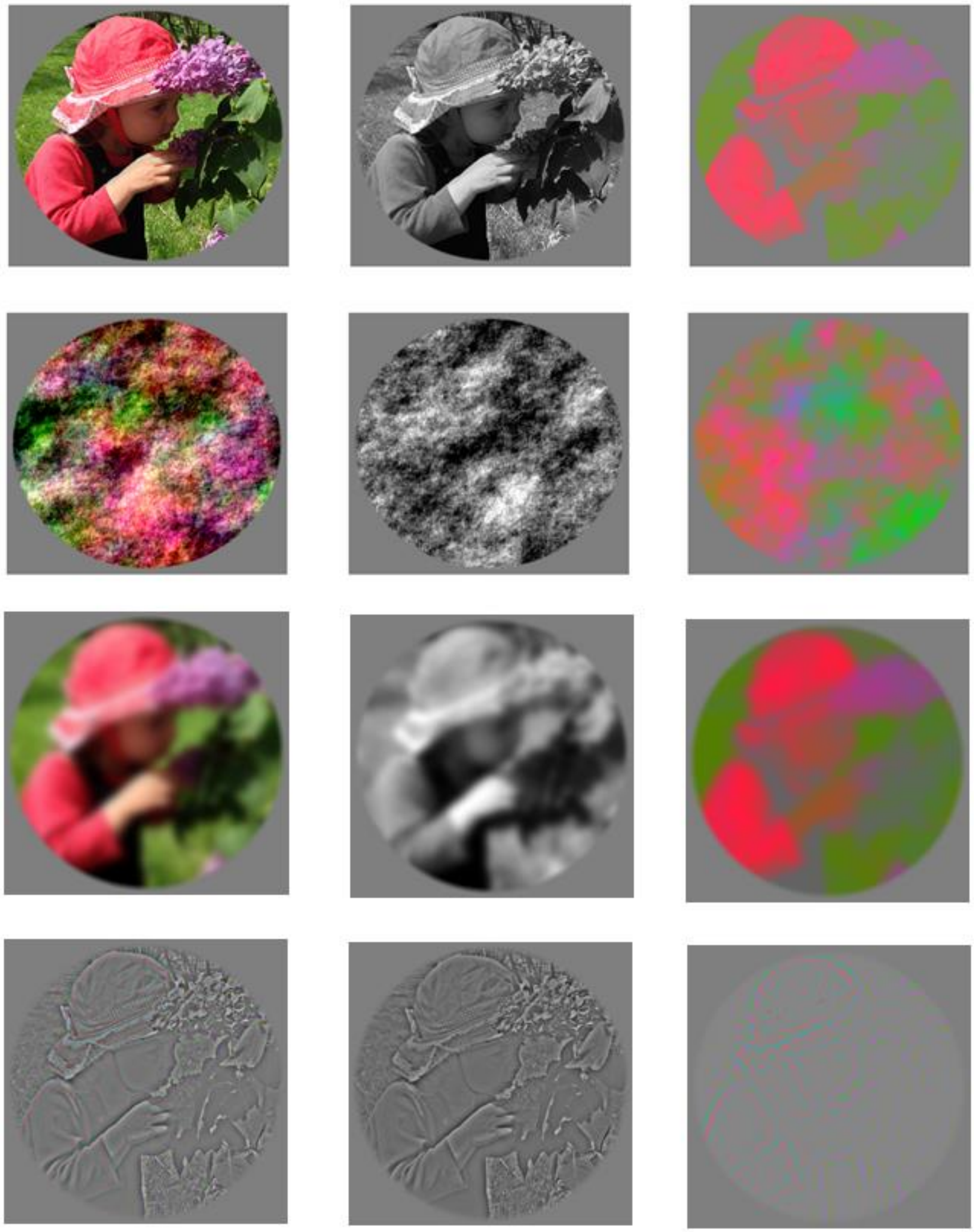

Fig. 1.Columns, left to right: colour, luminance and isoluminant conditions. Rows, top to bottom: real, phase scrambled, low-frequency, high-frequency conditions.

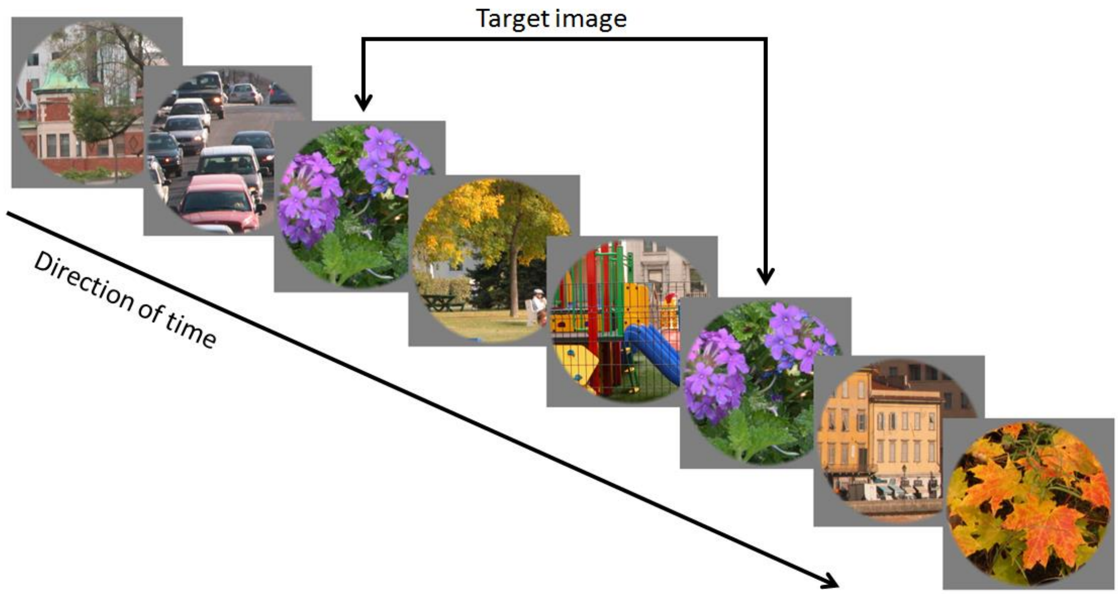

Fig. 2. A subsection of an example image stream is represented with a target scene present within it; the target scene is positioned with distractor level of 2.

As it was vital that individual observers were only exposed to each scene once (twice if it was a target) in order to avoid "false" false alarms, three different groups of observers took part in three experiments each of which contained different presentation times and numbers of distractor combinations. The groups and conditions were: (i) 1000 ms scene presentation time, with 0, 1 or 2 distractors (n=10), (ii) 1000 ms scene presentation time, with 3, 4 or 5 distractors (n=8), and (iii) 30 ms scene presentation time, with 3, 4 or 5 distractors (n=10). In total 5 blocks per condition were performed. An additional two groups of 9 participants took part in the spatial frequency conditions (one for the high- and one for the low-frequency condition).

Data analysis

Data was combined for each of the five blocks per condition per observer. The hit rate (H), defined as the proportion of target trials for which a correct response was given, was calculated. The false alarm rate (F), defined as the proportion of distractor images (including target images presented for the first time) for which an observer responded “seen before”. From these the bias free statistic d-prime (d’) was computed as the difference in z-scores between F and H (Eq. 3), for a first principles derivation of d’ see (Green and Swets, 1966). These d’ values were compared per distractor level for each condition in a series of repeated measures ANOVAs (see results section for full details).

$$d' = z(H) - z(F) \qquad \text{(Eq. 3)}$$

Results

Below we first summarize the data, and then provide two sets of statistical analyses. The first set comprises analyses of performance for the different presentation times and number of distractors, using repeated measures 3 x 3 ANOVAs, with factors *scene type* (i.e., colour, luminance or isoluminant) and *number of distractors* (0, 1, 2 or 3, 4, 5 depending on the subgroup being tested). F-statistic degrees of freedom are provided with an applied Greenhouse-Geisser corrections were appropriate, i.e., when there was a violation of sphericity as indicated via Mauchly’s test. Effect sizes (η^2) are reported for each ANOVA. All subsequent post hoc t-tests are reported with 2-tailed p-values, Bonferroni corrected for multiple comparisons. The second set of analyses is for the high- and low-spatial frequency band data, with ANOVAs for each of the colour, luminance and isoluminant conditions, with factors of *frequency* and *number of distractors*.

Effect of number of distractors and presentation time

For real scenes with 1000 ms presentation times and no distractors, performance is equal for all colour and luminance conditions. With the addition of more distractors performance decreases, the isoluminant condition more rapidly than the colour and luminance conditions (see Fig. 3a). This pattern of performance is also found with the phase scrambled scenes (see Fig. 3b). When the number of distractors is increased to 3, 4 and 5 (keeping presentation time at 1000 ms), there are again no differences between the colour and luminance conditions; however this time performance with the isoluminant scenes is worse (see Fig. 3c). For the phase scrambled scenes with distractor levels of 3, 4 and 5, performance in all conditions was at chance (d' ~ 0). Finally, with 3, 4 or 5 distractors, but a fast (30 ms) presentation time, a difference emerged between the full colour and luminance isolating scenes (Fig. 3d). Again, performance was at chance for the phase scrambled images. For a complete statistical analysis see below.

1000 ms presentation time with 0, 1 or 2 distractors

Real scenes: Significant main effects of scene type and number of distractors were revealed ($F(1.14, 10.26) = 16.72$, $p<.002$ ($\eta^2 = .65$) and $F(2, 18) = 26.55$, $p<.001$ ($\eta^2 = .75$), respectively). These effects were additionally qualified by a significant interaction ($F(1.96, 17.67) = 11.86$, $p<.001$ ($\eta^2 = .55$)).

Post hoc t-tests revealed no differences in d' for zeros distractors, with one distractor performance for full colour and luminance isolating scenes was significantly higher than for isoluminant scenes ($t(9)=8.72$, $p<.001$ and $t(9)= 6.38$, $p=.001$, respectively), this was also the case with two distractors (full colour vs. isoluminant $t(9)=5.04$, $p=.006$ and luminance isolating vs. isoluminant $t(9)= 8.18$, $p<.001$).

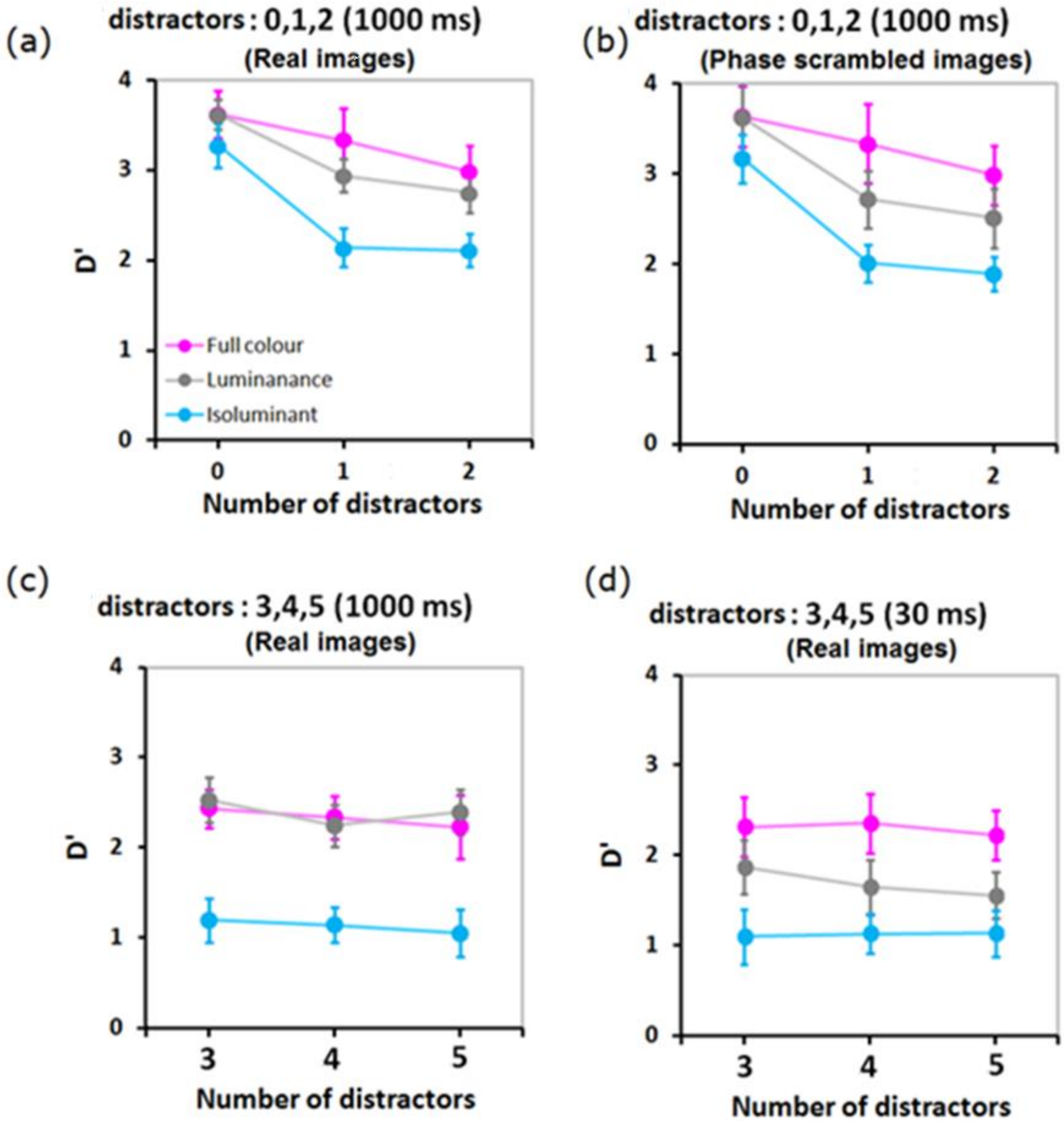

Fig. 3a-d. *All plots show performance as a function of distractor number for the colour (magenta data points), luminance (gray data-points) and isoluminant (blue data points) stimuli. (a) real scenes with a 1000 ms presentation time and 0, 1 or 2 distractors, (b) phase scrambled scenes with a 1000 ms presentation time and 0, 1 or 2 distractors, (c) real scenes with a 1000 ms presentation time and 3, 4 or 5 distractors and (d) real scenes with a 30 ms presentation time and 3, 4 or 5 distractors. Error bars represent ±2SE.*

Phase scrambled scenes: Significant main effects of scene type and number of distractors were revealed ($F(1.06, 9.51) = 15.35$, $p<.001$ and $F(1.25, 11.29) = 56.50$, $p<.001$, respectively). These effects were additionally qualified by a significant interaction ($F(1.67, 15.06) = 5.89$, $p<.016$).

Post hoc t-tests revealed no differences in d' for zeros distractors, with one distractor performance for full colour and luminance isolating scenes was significantly higher than for isoluminant scenes ($t(9)=11.24$, $p<.001$ and $t(9)= 4.72$, $p<.001$, respectively), this was also the case with two distractors (full colour vs. isoluminant $t(9)=5.84$, $p<.001$ and luminance isolating vs. isoluminant $t(9)= 19.70$, $p<.001$).

Real vs. Phase scrambled scenes: Additional t-tests revealed no differences between real and phase scrambled conditions for full colour, luminance isolating and isoluminant images, at each distractor level (t-values all in range: $-1.50 \leq t \leq 1.29$, p-values in range: $.17 \leq p \leq .96$).

1000 ms presentation time with 3, 4 or 5 distractors

Real scenes: A significant main effect of scene type but not number of distractors was revealed ($F(2, 12) = 16.20$, $p<.001$ ($\eta^2 = .73$) and $F(1.1, 6.6) = 0.57$, $p<.49$ ($\eta^2 = .086$)), respectively). No significant interaction was revealed ($F(1.579, 9.471) = 0.61$, $p<.53$ ($\eta^2 = .092$)).

Post hoc t-tests revealed a significant differences between full colour and isoluminant scenes at each distractor level (3: $t(6)=4.61$, $p=.033$, 4: $t(6)=3.75$, $p=.049$, and 5: $t(6)=4.22$, $p=.036$). Additionally, significant differences between full colour and isoluminant scenes at each distractor level (3: $t(6)=5.58$, $p=.013$, 4: $t(6)=4.10$, $p=.022$. and 5: $t(6)=4.89$, $p=.025$). No difference existed between full colour and luminance isolating scenes (t-values all in range: $-0.64 \leq t \leq 0.53$, p-values: $p=1$).

Phase scrambled scenes: For this load level, i.e., 3, 4, or 5 distractors, the task was too difficult and hence no useable data was collected.

30 ms presentation time with 3, 4 or 5 distractors

Real scenes: A significant main effect of scene type but not number of distractors was revealed ($F(2, 18) = 31.27$, $p<.001$ ($\eta^2 = .78$) and $F(2, 18) = 2.89$, $p<.082$ ($\eta^2 = .24$)), respectively). No significant interaction was revealed ($F(4, 69) = 1.98$, $p=.12$ ($\eta^2 = .18$)).

Post hoc t-tests revealed a significant differences between full colour and luminance isolating conditions for every distractor level (3: $t(9)=4.70$, $p=.010$, 4: $t(9)=2.46$, $p=.014$ and 5: $t(9)=3.61$, $p=.046$). Significant differences also existed for between the full colour and isoluminant conditions for every distractor level (3: $t(9)=5.64$, $p<.001$, 4: $t(9)=6.74$, $p<.001$ and 5: $t(9)=6.27$, $p<.001$). Finally, significant differences between the luminance isolating and isoluminant conditions were found for distractor levels 3 and 4 ($t(9)=5.14$, $p=.005$ and $t(9)=1.95$, $p=.043$, respectively). No significant difference was found for the highest distractor level of 6 ($t(9)=2.07$, $p=.66$).

Phase scrambled scenes: For this load level, i.e., 3, 4, or 5 distractors (coupled with fast presentation times), the task was too demanding and no useable data was collected.

Effect of removing high and low spatial frequencies

Fig. 4a-b plots performance for both low and high frequency filtered scenes as a function of distractor number for each condition. The pattern of results was similar for colour, luminance and isoluminant conditions: no significant effect of spatial frequency was found. However, there was an effect of the number of distractors; there was a difference between the no distractors and either 1 or 2 distractor conditions. For a detailed analysis see below.

Spatial frequency and colour stimuli

No significant main effect of spatial frequency was found while a significant main effect for the number of distractors was revealed ($F(1, 7) = 0.71$, $p=.43$ ($\eta^2 = .093$) and $F(2, 14) = 37.56$, $p<.001$ ($\eta^2 = .84$), respectively). No significant interaction was revealed ($F(1.2, 8.50) = 0.18$, $p<.73$ ($\eta^2 = .025$)).

Post-hoc t-tests revealed that this was due to differences in performance for no distractors vs. either 1 or 2 distractors, this was that case for both high and low frequency scene types (High frequency: 0 vs. 1 distractors $t(7)=5.17$, $p<.008$, 0 vs. 2 distractors $t(7)=10.67$, $p<.001$, 1 vs. 2 distractors $t(7)=1.92$, $p=.58$. Low frequency: 0 vs. 1 distractors $t(7)=3.88$, $p=.036$, 0 vs. 2 distractors $t(7)=7.75$, $p=.001$, 1 vs. 2 $t(7)=1.44$, $p=1$).

Spatial frequency and Luminance isolating stimuli

No significant main effect of spatial frequency was found while a significant main effect for the number of distractors was revealed ($F(1, 7) = 0.71$, $p=.48$ ($\eta^2 = .074$) and $F(2, 14) = 29.45$, $p<.001$ ($\eta^2 = .81$), respectively). No significant interaction was revealed ($F(2, 14) = 0.18$, $p<.74$ ($\eta^2 = .025$)).

Post-hoc t-tests revealed that this was due to differences in performance for no distractors vs. either 1 or 2 distractors, this was that case for both high and low frequency image types (High frequency: 0 vs. 1 distractors $t(7)=6.10$, $p<.001$, 0 vs. 2 distractors $t(7)=4.88$, $p=.002$, 1 vs. 2 $t(7)=1.61$, $p=.91$. Low frequency: 0 vs. 1 distractors $t(7)=4.77$, $p=.002$, 0 vs. 2 distractors $t(7)=5.08$, $p=.009$, 1 vs. 2 $t(7)=2.56$, $p=.23$).

Spatial frequency and isoluminant stimuli

No significant main effect of spatial frequency was found while a significant main effect for the number of distractors was revealed ($F(1, 7) = 0.034$, $p=.86$ ($\eta^2 = .005$) and $F(2, 14) = 48.00$, $p<.001$ ($\eta^2 = .87$), respectively). A significant interaction was revealed ($F(2, 14) = 0.013$, $p<.013$ ($\eta^2 = .46$)).

Post-hoc t-tests revealed that this was due to differences in performance for no distractors vs. either 1 or 2 distractors, this was that case for both high and low frequency image types (High frequency: 0 vs. 1 distractors $t(7)=5.17$, $p=.008$, 0 vs. 2 distractors $t(7)=10.67$, $p<.001$, 1 vs. 2 $t(7)=1.92$, $p=.58$. Low frequency: 0 vs. 1 distractors $t(7)=3.89$, $p=.036$, 0 vs. 2 distractors $t(7)=7.75$, $p<.001$, 1 vs. 2 $t(7)=1.44$, $p=1$).

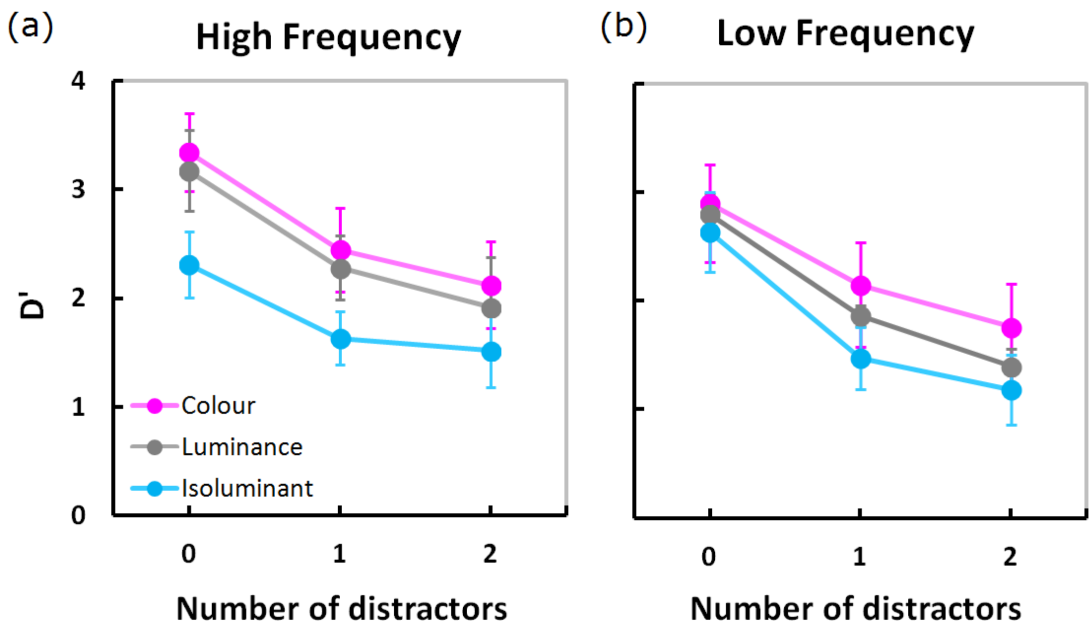

Fig. 4a-b. *Both plots show performance as a function of distractor number for the colour (magenta data points), luminance (gray data-points) and isoluminant (blue data points) stimuli. Panel (a) plots the low frequency, while (b) plots the high frequency isolating scenes, both presented with a 1000 ms presentation time and 0, 1 or 2 distractors. Error bars represent ±2SE*

Effect of spatial frequency within conditions

Between the conditions of the low spatial frequency band t-tests revealed no differences between the colour, luminance or isoluminant image types, for any distractor level (all t-value in range, $0.18 \leq t \leq 5.33$ values in range: $.085 \leq p \leq 1$).

Between the high frequency conditions the data indicated differences in performance for distractor levels of zero and one between full colour vs. isoluminant ($t(7)=3.52$, $p=.01$ and $t(7)=2.83$, $p=.025$, respectively) and distractor levels of zero and one between luminance only vs. isoluminant scenes ($t(7)=5.33$, $p=.001$ and $t(7)=2.48$, $p=.042$, respectively). No differences were found between conditions when 2 distractors where present (t-value in range, $0.87 \leq t \leq 2.17$ values in range: $.067 \leq p \leq .41$).

Discussion

The following summarises the main findings of the study.

(i) For 1000 ms presentation times and no distractor images VWM performance is equal for colour, luminance and isoluminant scenes, irrespective of whether real or phase scrambled.

(ii) For 1000 ms presentation times with 1 or 2 distractors performance for colour and luminance stimuli is equal, while performance for isoluminant scenes is lower, a pattern similar for both real and phase scrambled images.

(iii) For 1000 ms presentation times with 3, 4 or 5 distractors there is again no difference between colour and luminance scenes, with worse performance for isoluminant scene. However this time the task was impossible with phase scrambled scenes.

(iv) For brief 30 ms presentation times and 3, 4 or 5 distractors, a difference is observed between colour and luminance scenes, with higher performance for colour scenes. Again, performance was impossible for the phase scrambled images.

(v) For stimuli containing only a low or high frequency component performance declined with the number of distractors present with colour, luminance and isoluminant scenes. However there was no difference between colour, luminance or isoluminant at any distractor level.

Given a long enough exposure time VWM performance for colour and luminance defined images of real natural scenes is equal at all distractor levels. Only at very brief exposure times does an advantage for colour over luminance scenes emerge. This is consistent with data presented by Gegenfurtner and Rieger (2000), they found that during a match-to-sample task target scenes were easily identified given long presentation times. However, for briefly presented stimuli an advantage emerged for full colour over greyscale scenes. Recall performance for isoluminant scenes is invariably worse.

It is clearly important for VWM to have intact scene structure during encoding, as the task is rendered impossible when phase scrambled scenes are presented with high distractor numbers and/or short exposure times. Performance for the isoluminant condition was lowest, perhaps a reduced ability to extract scene structure from only isoluminant information accounts for the relatively poor performance observed in this condition.

Overall the data reflects VWM's preference for utilising scene structure when there is sufficient time to do so, but when exposure time is restricted and complex scene structure cannot be processed and stored in VWM, other low-level properties of the scene, i.e. its chromatic content, are encoded and maintained over the short-term time scales of VWM.

Acknowledgements

This research is supported by Canadian Institute of Health Research grant #MOP 123349 given to F.K.

References

Cowan, Nelson. 2000. "The Magical Number 4 in Short-Term Memory : A Reconsideration of Mental Storage Capacity." (4):87–185.

Daneman, M; and P. A. Carpenter. 1980. "Individual Differences in Working Memory and Reading." 466:450–66.

Gegenfurtner and Rieger. 2000. "Sensory and cognitive contributions of color to the recognition of natural scenes." Current Biology, 10(13):805-808.

Green, D. M. and J. A. Swets. 1966. *Signal Detection Theory and Psychophysics*. Wiley.

Harding, Glen and Marina Bloj. 2017. “Real and Predicted in Fl Uence of Image Manipulations on Eye Movements during Scene Recognition.” 10(2010):1–17.

Isola, Phillip, Jianxiong Xiao, Devi Parikh, Antonio Torralba, and Aude Oliva. 2014. “What Makes a Photograph Memorable?” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36(7):1469–82.

Kirchner, W. K. (1958). "Age differences in short-term retention of rapidly changing information". Journal of Experimental Psychology. ***55*** *(4): 352–358. doi:10.1037/h0043688.*

Kosilo, M., C. Haenschel, and J. Martinoivc. 2015. “Preferential Inputs of Luminance Signals for Visual Working Memory.” *Perception* 44:9–10.

Magnussen, S. and S. Dyrnes. 1994. “High-Fidelity Perceptual Long-Term Memory.” *Psychological Science* 5(2):99–102.

Marx, Svenja and Onno Hansen-goos. 2017. “Rapid Serial Processing of Natural Scenes : Color Modulates Detection but Neither Recognition nor the Attentional Blink Wolfgang Einh a.” 14(2014):1–18.

Miller, G. A. (1956). "The magical number seven, plus or minus two: Some limits on our capacity for processing information". Psychological Review. 63 (2).

Miyake, A.; Shah, P., eds. (1999). Models of working memory. Mechanisms of active maintenance and executive control. Cambridge University Press

Oliva, Aude and Philippe G. Schyns. 2000. “Diagnostic Colors Mediate Scene Recognition.” 210:176–210.

Olmos, Adriana and Frederick A. A. Kingdom. 2004. “A Biologically Inspired Algorithm for the Recovery of Shading and Reflectance Images.” 33.

Potter, M. C. 1976. “Short-Term Conceptual Memory for Pictures.” *Journal of Experimental Psychology: Human Learning, Memory* 2(5):509–22.

Spence, Ian, Patrick Wong, Maria Rusan, and Naghmeh Rastegar. 2006. “How Color Enhances Visual Memory for Natural Scenes.” 17(1):1–6.

Wenger, Michael J. and James T. Townsend. 2000. “Spatial Frequencies in Short-Term Memory for Faces : A Test of Three Frequency-Dependent Hypotheses.” 28(1):125–42.

Wichmann, Felix A., Lindsay T. Sharpe, and Karl R. Gegenfurtner. 2002. “The Contributions of Color to Recognition Memory for Natural Scenes.” 28(3):509–20.

Yoonessi, Ali, Frederick A. A. Kingdom, and Samih Alqawlaq. 2008. “Is Color Patchy ?” 25(6):1330–38.