Colour and Object Representation: Insights into Human Visual Processing

Sana Hassan Mansour
Faculty of Medical Sciences, Tripoli
www.mstct.edu.ly
sanamansour6@gmail.com

Abstract

This study investigates the role of colour and texture in object recognition, addressing gaps in previous research that focused on unnatural objects. Using behavioural tasks and fMRI, we explored how natural chromatic texture influences object identification. A backward masking paradigm with varying Stimulus Onset Asynchronies (SOAs) of 40 ms, 100 ms, and 250 ms was employed to examine visual processing timing. The results show that colour significantly improves object recognition performance, more so than texture. The varying SOA results suggest that texture and colour processing take longer than shape processing. fMRI scans identified two key areas in the visual cortex: the Collateral Sulcus (CoS) for colour processing and the Lateral Occipital Complex (LOC) for object recognition. These findings enhance our understanding of how the brain integrates colour and texture information during natural object recognition, contributing to the broader knowledge of visual perception mechanisms.

Introduction

Object recognition is an essential function that enables us to interact with the surrounding environment. It can be vital for survival, such as in recognizing and identifying a poisonous wild mushroom before deciding whether to eat it or not. Recent research continues to investigate how the visual system enables object recognition, yet findings remain complex and sometimes contradictory. Studies suggest that the visual system processes information across multiple dimensions—such as color, luminance, motion, and depth—integrating these features at higher cognitive levels to form a cohesive perception of objects (Ge et al., 2022). While earlier research primarily focused on geometric aspects like size and shape, recent studies have begun to emphasize the importance of surface properties, including color and texture, in object recognition. For instance, the integration of color and motion cues has been shown to significantly impact depth perception and object discrimination, highlighting the multifaceted nature of visual processing (Scabini et al., 2024).Consequently, few studies have examined the role of colour and texture in object recognition, especially for natural objects. Natural objects such as fruits and vegetables often possess surface variations in colour (chromatic texture), which may be due to intrinsic patterns (e.g., the yellow, green, and brown spots on a banana skin) or to 3D shape or surface structure (e.g., the pockmarks on a lemon or strawberry). Recent studies have further refined our understanding of the visual processing pathways. When light enters the eye, it is detected by photoreceptors in the retina, and this information is transmitted via retinal ganglion cell axons through the optic nerve to the lateral geniculate nucleus (LGN) in the thalamus. From the LGN, signals are relayed to the primary visual cortex (V1), where initial visual processing occurs (Zhang et al., 2024). Within the retinal ganglion layer, two primary types of cells are recognized: parvocellular cells, which specialize in high-resolution form and color processing, forming the basis of the ‘what’ pathway, and magnocellular cells, which process motion and depth, contributing to the 'where' pathway (Choi et al., 2023). The anatomical division of these pathways is evident in the LGN, where the layers receiving parvocellular and magnocellular input are functionally distinct (Martinez et al., 2024). From V1, visual information diverges into two primary processing streams: The ventral stream, projecting to V4 and the inferior temporal cortex (IT), is crucial for object recognition, color processing, and scene understanding (Zhang et al., 2024). The dorsal stream, projecting to the middle temporal area (MT or V5) and the medial superior temporal area (MST), is responsible for processing motion, spatial relationships, and visually guided actions (Martinez et al., 2024).

Methods and Materials:

Experiment 1:Behavioural Task:

The behavioral experiment utilized Matlab 7.1, with visual stimuli presented on a calibrated CRT monitor in a darkened room, ensuring optimal viewing conditions. Monitor calibration is critical to minimize variations over time.

Participants

Twelve volunteers (4 males and 8 females; mean age 24 years, range 18-30) with normal colour vision participated. All subjects provided informed consent and underwent color vision tests, including the Neitz test, Ishihara Test, and Farnsworth-Munsell 100-hue test, conducted under artificial lighting (lamp D65) to mimic daylight conditions.

Stimuli and methods

The stimuli used were based on an image database from medical school at Newcastle university lab using a colorimetrically calibrated digital camera. The database contained images of natural objects (fruits and vegetables). The natural objects were divided into 3 sets (Set 1: banana, courgette and carrot; Set 2: potato, lime and carrot; Set 3: strawberry, kiwi and carrot). The carrot was repeated in each set to serve as a control for set effects. Using the colorimetric image data and customized image processing programmes.

The information about colour, texture and shape was systematically manipulated into 8 different conditions, (combined shape and texture cues), (texture cue alone), (colour cue alone), (combined colour and texture cues), (combined colour and shape cues), (Shape alone), (Shape alone with high luminance contrast),(combined colour, texture and shape cues).

Backward masking:

To explore the timing of visual processing, a backward masking paradigm was implemented using a random dot chromatic mask. The mask interfered with stimulus visibility, presented at three stimulus onset asynchronies (SOAs): 40ms, 100ms, and 250ms. Masks were sized according to the stimulus dimensions, with the whole image mask measuring 8cm × 3cm and patch images measuring 5cm × 5cm.

The behavioural experiment procedure and trial timing:

The experiment lasted approximately 2 hours and 45 minutes in a darkened room, with subjects seated 57cm from the monitor. The CRT monitor was warmed up for 30 minutes before testing. Each trial commenced with a 267ms fixation period, during which subjects focused on the center of the screen. This was followed by a 27ms stimulus presentation and then the mask, based on the SOA. A total of 72 blocks were created, comprising three sets and eight conditions across three SOAs, resulting in 45 trials per block. To minimize after-image effects, each image was slightly repositioned on the screen (figure 1).

Subjects received detailed instructions at the beginning of the experiment, followed by a practice block. In each trial, subjects pressed keys corresponding to the displayed images as quickly as possible. Specific instructions for each block (e.g., “Press Z for Banana, V for Carrot, and M for Courgette”) were presented on the screen. Data on reaction times and accuracy were collected and analyzed using Excel and SPSS.

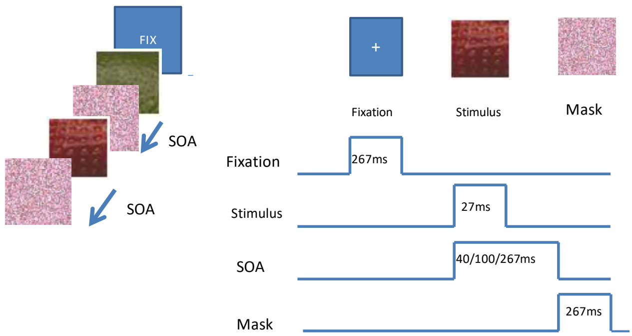


Figure 1: Behavioral task trial timing.

<u>Experiment 2: fMRI (functional Magnetic Resonance Imaging) study:</u>

This part of the experiment has 2 steps.

Experiment 2A: Two volunteers with normal color vision—one male (27 years old) and one female (32 years old)—were recruited for this phase. The session began with a 4-minute structural scan, followed by a 6-minute localizer scan to identify areas of interest prior to the main fMRI experiment.

Experiment 2B: This phase focused on event-related adaptation. The same two subjects participated in a subsequent scanning session where they viewed images of both natural and nonsense objects, which varied in color, texture, and shape cues.

Experiment 2.A: Localiser scan Stimuls

To localize "object" and "color" areas, four types of images were used:

- Object Area: Whole images of identifiable objects (e.g., guitar, rubber duck) and scrambled versions that retained color and contrast but rendered the objects unrecognizable.
- Color Area: Colored Mondrian patches (high contrast) and grayscale Mondrians, each sized 400x400 pixels.

Experiment: 2.B: Event-related adapation

This phase involved 12 original object images categorized into three groups: Textured-3D (natural chromatic texture and shading), uniform-flat (no shading or texture) and control (nonsense objects).

Functional data acquisition

Functional T2-weighted images were acquired using a 3T Philips Intera Achieva MRI scanner, with parameters optimized for image quality and resolution.

MRI Data pre-processing

Data analysis utilized SPM5 (Statistical Parametric Mapping), following several pre-processing steps:

1. Realignment of scans to correct for motion.
2. Slice timing adjustment for event-related data.
3. Normalization to MNI space.
4. Smoothing with a 6mm Gaussian Kernel.

Results

Behavioral task result:

Accuracy

The average accuracy rates of the 12 subjects under the 8 different conditions (shape with original luminance, shape with high luminance, colour, texture, shape and colour, shape and texture, colour and texture, and finally, colour, texture and shape) with the 3 SOAs, SOA 40ms, SOA 100ms and SOA 250ms are illustrated in (figure 2). The accuracy across the average of the 3 SOAs recorded that SOA40 (mean: 0.865) was significantly less than SOA100 (mean: 0.929) and SOA250 (mean: 0.928), while there was no significant difference between SOA100 and SOA250. An Anova test showed a significant difference between SOA 40 and SOA 100, $F(2,254) = 3.812$, $p = 0.023$ and SOA 40 and SOA 250, $F(2,233) = 7.335$, $p = 0.001$. The accuracy rate over the 3 SOAs was worse with texture only than all other conditions. See figure 2 and table 1 (the average score of the 8 conditions at the 3 different SOAs, 40, 100 and 250).

Accuracy at SOA40: The accuracy was worse with texture alone (mean: 0.563) than all other conditions. Post hoc tests revealed that accuracy for the texture alone condition was less compared to all other conditions ($p<.001$), while the mean of correct responses was more with the whole image condition (mean: 0.958) than all other conditions. There was no significant difference between most of all other conditions.

Accuracy at SOA 100: At SOA 100, the accuracy for the texture alone condition was significantly improved (mean: 0.805, $p < 0.05$) compared to texture only at SOA 40, however, it was still significantly less than for all other conditions. Post hoc tests revealed that for the texture alone condition accuracy was significantly worse than for all other conditions ($p<.001$). There was no significant difference between all other conditions.

Accuracy at SOA 250: The accuracy level with the texture alone condition was slightly better at SOA 250 (mean: 0.864) than that at SOA 100 (mean: 0.805), but it was still significantly less than all other conditions. Post hoc tests revealed that texture alone is significantly worse than all other conditions ($p<.001$). The shape alone condition (original luminance) at SOA 250 was significantly less accurate than shape and texture, shape and colour and whole image conditions ($p < 0.05$).

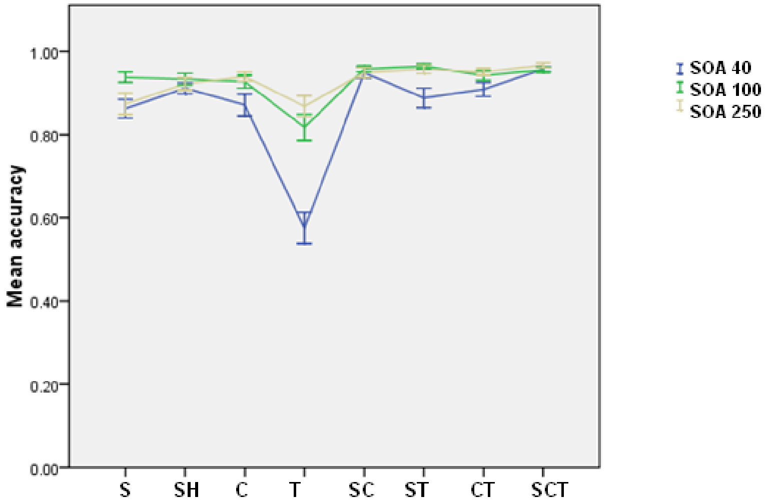


Figure 2: The graph illustrates mean accuracy as a function of the 8 different conditions (S=shape with original luminance, SH= shape with high luminance, C= colour, T= texture, SC=shape and colour, ST= shape and texture, CT= colour and texture, and finally, SCT= colour, texture and shape with the 3 SOAs, SOA 40ms, SOA 100ms and SOA 250ms. Error bars: +/- 1 SE

SOA	T	ST	TC	C	SC	S	SH	TSC	Total
40	56.3%	87%	90.2%	86.5%	94.9%	85.4%	92.1%	95.8%	86%
100	80.5%	95.6%	91.3%	93.1%	95.5%	88.8%	92.1%	95.9%	91.6%
250	86.4%	95.8%	95.4%	92.4%	94.6%	88.8%	90.7%	96.5%	92.5%
Total	74.4%	92.8%	92.3%	90.6%	95%	87.6%	91.6%	96%	74.4%

Table 1: Gives information about accuracy means over the 8 conditions at the 3 SOAs. T= texture, ST= shape and colour, TC= texture and colour, C = colour, SC = shape and colour, S= shape with original luminance, SH= shape with high luminance and TSC = whole image (texture, shape and colour)

Reaction time

A statistical analysis for the reaction time (response time) was run for only accurate trials. An Anova test revealed a significant effect of variation of mask presenting time on the reaction time for all conditions, ($p< 0.05$), $F (7,573) = 10.180$. A Tukey HSD test showed that the total average reaction time of "colour and texture" together is significantly quicker than texture alone ($p= 0.000$) and colour alone ($p= 0.043$).

Reaction time at SOA 40:

Post hoc tests revealed that the reaction time for colour alone (mean: 697.2; sd: 623.4) was significantly longer than all other conditions

($p< 0.05$). Texture alone (mean: 671.6; sd: 470.61) was significantly longer than the two-cue conditions with shape, "shape and colour", "shape and texture", and to whole image ($p < 0.05$).

Reaction time at SOA 100

Post hoc tests showed that the reaction time for texture alone and colour alone were also longer than the di-cue conditions, "shape and colour", "colour and texture", and to whole image as well as to shape alone conditions (with original luminance and high luminance), ($p< 0.05$).

Reaction time at SOA 250:

Colour alone, "colour and texture and shape" were significantly faster than shape alone condition ($p <0.05$).

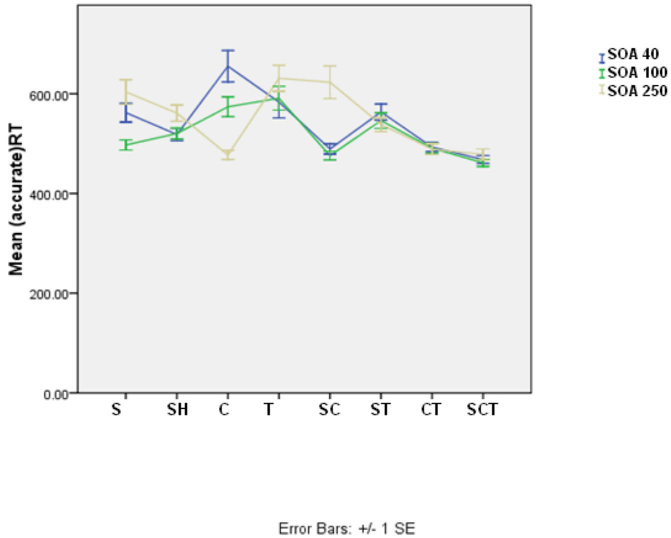


Figure 3: The graph illustrates mean reaction time as a function of the 8 different conditions (S= shape with original luminance, SH= shape with high luminance, C= colour, T= texture, SC=shape and colour, ST=shape and texture, CT=colour and texture, and finally SCT= colour, texture and shape with the 3 SOAs, SOA 40ms, SOA 100ms and SOA 250ms.

Performance

The performance (accuracy/reaction time) was calculated to include the effect of accuracy as well as reaction time. A Tukey test showed that the performance across the average of all SOAs was significantly improved by increasing the mask presentation time from 40 ms (mean: 1.72) to 100 ms (mean: 1.89), p= 0.003, while there was no significant difference between presenting the mask at 100 ms or 250 ms (mean: 1.83), p = 0.441.

Performance at SOA 40:

Texture alone performance (mean: 1.16) was worse than all other conditions (p< 0.001). The performance with colour alone (mean: 1.49) was significantly less than shape with high luminance (mean: 1.82), the two-cues “shape and colour”, “colour and texture” and whole image. “Shape and colour” performance was significantly better than shape alone (with original luminance) and colour alone, texture alone and “shape and texture”, $p < 0.05$. The performance of colour and texture together is better than the performance in colour alone and texture alone.

Performance at SOA 100:

A Tukey test revealed that the performance of texture alone condition at SOA 100 was significantly improved compared to performance of texture alone at SOA 40 ($p = 0.004$). However, it was still significantly worse than all other conditions, including colour alone and shape alone. Colour alone performance was also significantly increased (mean: 1.75) compared to that of colour only at SOA 40 ($p = 0.006$), but it was significantly less than “shape and colour” and “colour and texture”. The performance of shape alone (original luminance) was significantly increased (mean: 1.93) compared to that at SOA 40 ($p=0.004$). Performance of “shape and colour” (mean: 2.06) was significantly more than” shape and texture” (mean: 1.84).

Performance at SOA 250:

A Tukey test showed no significant difference between the performance of texture alone at SOA 100 and texture alone at SOA 250 ($p = 0.969$) which was significantly worse than colour alone, “shape and colour”, “colour and texture” and whole image. Colour alone performance was significantly improved at SOA 250, compared to that at SOA 40 ($p=0.000$) and SOA 100 ($p = 0.006$), and significantly better than performance for shape alone (original luminance) and texture alone; in addition, the performance of colour alone was the same as whole image and colour and texture ($p > 0.05$). The performance of shape alone (original luminance) at SOA 250 was significantly less compared to that at SOA 100 (p = 0.00).

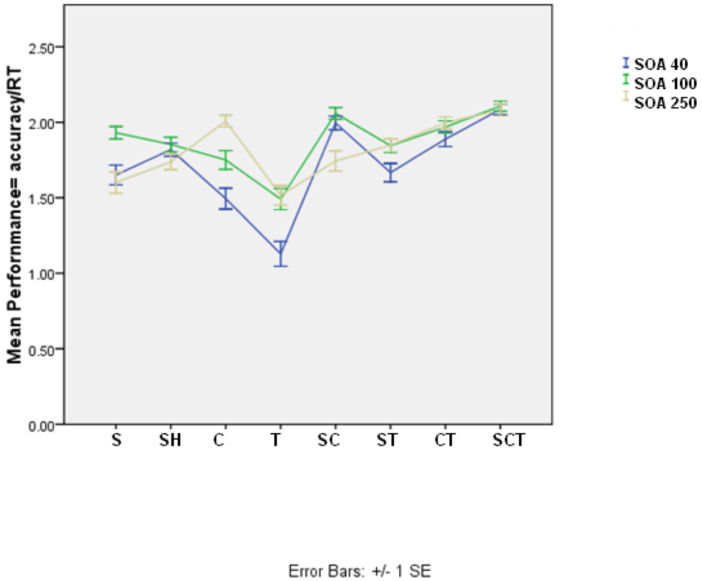


Figure 4: The graph illustrates mean performance as a function of the 8 different conditions (S= shape with original luminance, SH= shape with high luminance, C= colour, T= texture, SC= shape and colour, ST= shape and texture, CT= colour and texture, and finally, SCT= colour, texture and shape with the 3 SOAs, SOA 40ms, SOA 100ms and SOA 250ms.

Table 2 : Shows all the mean performance scores for every condition over the 3 SOAs. S=Shape with original luminance, SH= shape high lum, C = colour, T = texture, SC = shape colour, ST = shape texture, CT = colour texture, and SCT = shape colour and texture.

	S	SH	C	T	SC	ST	CT	SCT
SOA40	1.65	1.82	1.50	1.16	2	1.67	1.89	2.08
SOA100	1.93	1.85	1.75	1.56	2.06	1.84	1.97	2.11
SOA250	1.65	1.74	2.01	1.53	1.74	1.85	2	2.08

Colour, texture and shape performance summary over SOA 40, 100 and 250 ms:

<u>**Shape**</u>: the performance of shape alone was significantly more than colour and texture over SOA 40 and 100; and the performance of shape peaked at SOA 100.

<u>**Colour**</u>: the performance of colour was significantly more than texture over the 3 SOAs, but it was significantly less than the shape condition at SOA 40 and 100, it was only significantly more than shape at SOA 250 and the same as the performance of whole image and colour and texture.

<u>**Texture**</u>: the performance was worst for texture alone compared to colour alone (over the 3 SOAs) and shape alone (over SOA 40 and SOA 100). The performance of texture was significantly improved at SOA 100, while there was no significant difference between SOA 100 and SOA 250. See figure 5.

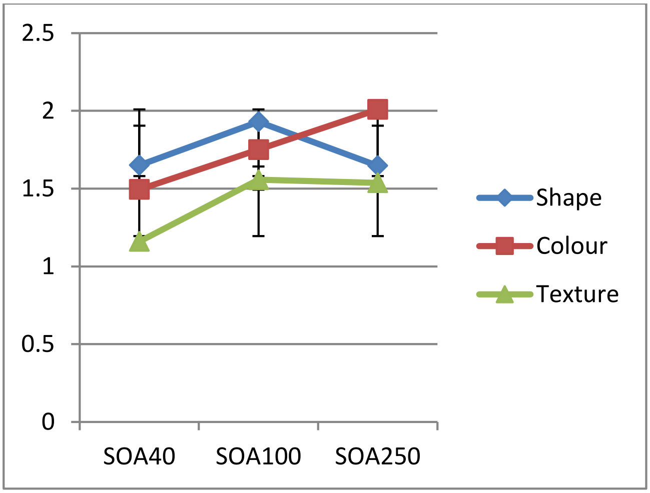


Figure 5: Colour, texture and shape performance summary over SOA 40, 100 and 250 ms

Colour, texture and shape cues and performance

increase**:**

To understand whether colour, texture or shape is most important in object identification, the effect of adding cues was examined; for example, to find out the effect of adding colour to texture, the performance of combined colour texture was divided by the performance of texture alone.

- If the result > 1 (adding colour to texture increases the performance).
- If the result = 1 (the performance of both “colour texture” and texture are the same).
- If the result < 1 (the performance of “colour texture“is less than texture alone.

Increase in the performance at SOA 40:

Adding colour to texture significantly improved the performance compared to adding colour to shape ($p = 0.012$) and adding shape to “colour and texture” ($p=0.002$). Adding texture did not significantly improve the performance compared to other cues; for example, adding texture to shape and to “shape and colour” was significantly less in terms of effect than adding colour to texture ($p< 0.005$). See figure 6.

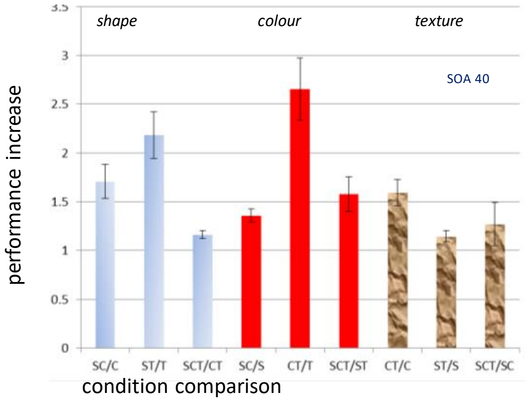


Figure 6: Colour, texture and shape cues effect on increase in performance. S = shape, C = colour, T = texture at SOA 40.

Increase in the performance at SOA 250:

The performance was significantly increased through adding colour to texture compared to adding shape to colour (p=0.01), adding shape to "colour and texture" (p=0.005) and adding texture to colour (p=0.002).

Increase in performance at SOA 100:

The figure was the same as at SOA 40. Adding colour to texture significantly improved the performance compared to adding colour to shape (p= 0.006) and adding shape to "colour and texture" (p= 0.008). Adding texture to shape was significantly less in terms of effect compared to adding colour to texture (p=0.000) and adding shape to texture (0.003), as well as adding shape to colour (p = 0.048). In addition, the increase in performance caused by adding texture to "shape and colour" was significantly less than that from adding colour to texture. See figure 7.

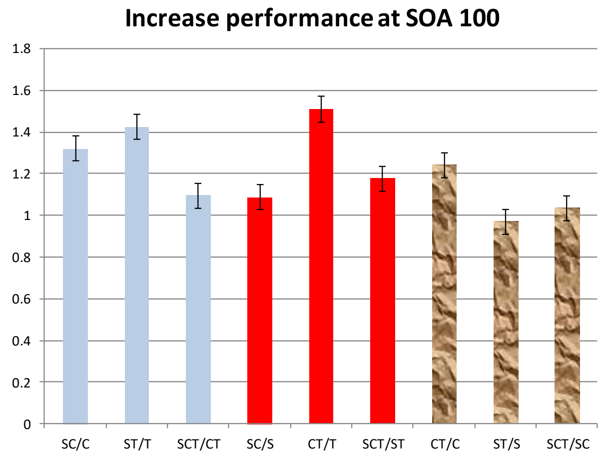


Figure 7: Colour, texture and shape cues effect on increase in performance. S = shape, C = colour, T = texture at SOA 100.

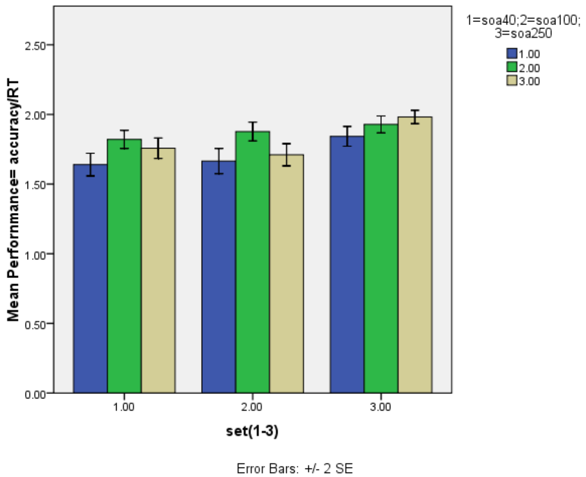
Figure 8: Colour, texture and shape cues effect on increase in performance. S= shape, C =colour, T= texture at SOA 250

Object and set effect

In the behavioral experiment, 3 sets of fruits and vegetables were used, the idea being to investigate whether using objects with highly diagnostic chromatic textures (for example, strawberry and kiwi), as in set 3, would increase the performance level more than in set 1 (carrot and courgette) and set 2 (lime and potato) which have approximately the same shape outline with less diagnostic chromatic texture (as suggested above). The carrot was used in every set as a control.

The performance in set 3 was significantly more than in set 1 over the 3 SOAs, SOA 40 (p = 0,001), SOA 100 (p = 0.001) and SOA 250 (p =

0.001). The performance for the carrot in set 1 was significantly less than for the carrot in set 3 (p = 0.003) and kiwi in set 3 (p = 0.001). The performance was significantly more for the kiwi than for the potato in set 2. See figure 9 and 10



.

Figure 9: Performance as a function of set.

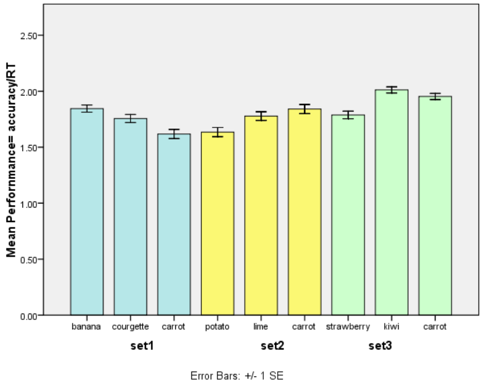


Figure 10: Performance as a function of object.

<u>Neuro-imaging experiment result</u>

Localiser results:

Whole brain statistical analysis was performed using SPM5. The statistical map was set on a threshold of K = 10 voxels (i.e. activity clusters larger than 10 voxels only are identified); the t-threshold varied between $p < 0.001$-0.0001. The results were very similar to the Talaraich coordinates reported by Cant and Goodale (2007) for both colour and object localiser.

<u>Object localiser</u>

A significant activation ($p < 0.000$) was recorded within the right and left lateral occipital area (LOC) or the "object area" (Cant and Goodale 2007; Kourtzi and Kanwisher 2000). Table 3 shows a comparison between Cant and Goodale LOC Talaraich coordinates and our results; figure 11 shows significant clusters within LOC.

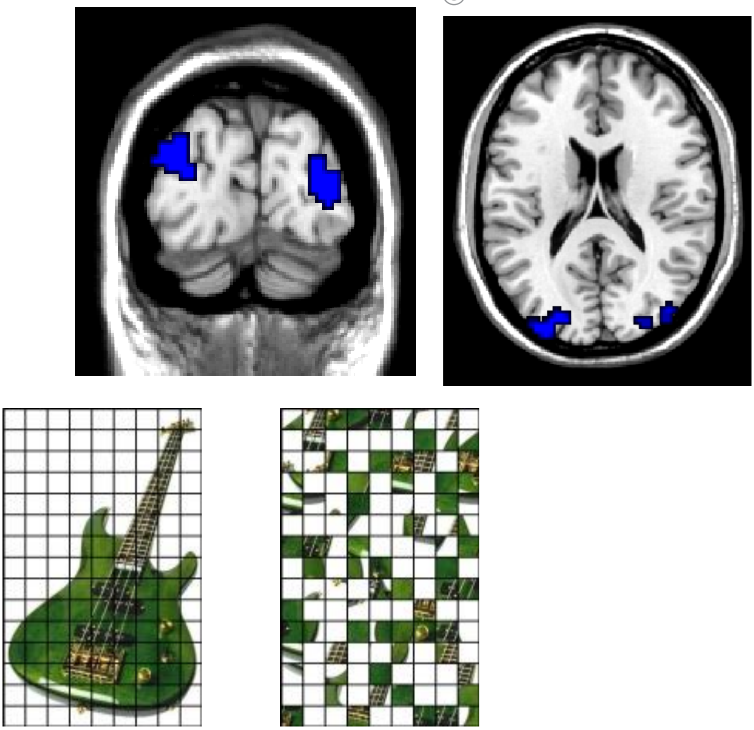


Figure 11: Brain slices (left: axial view; right: coronal view) show activity clusters within the object area (LOC or lateral occipital complex) of each hemisphere, identified via the object versus scrambled task (far right).

Table 3: Show a Comparison between locations of LOC in our study and Cant and Goodale (2007); Talaraich coordinates.

LOC	X	Y	Z	P value
Right LOC (Talaraich. C)	47.5	-76.4	5	0.000
Cant and Goodale LOC	36	-69	-6	
Left LOC (Talaraich. C)	- 47.5	- 81	9	0.000
Cant and Goodale LOC	- 50	- 69	- 4	

Colour localiser

Significant clusters ($p < 0.000$) were localised within anterior and posterior V4 areas that were respond better to coloured versus grey Mondrians. Anterior V4 area is the equivalent of Cant and Goodale's CoS and posterior V4 area is the equivalent of Cant and Goodale's IOG (see Table 4 and figure 12).

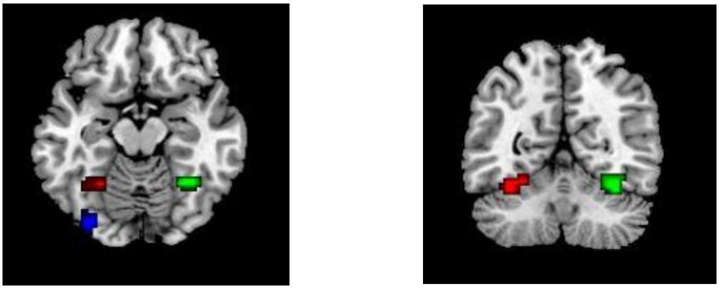


Figure 12: Brain slices show areas that respond better to coloured versus grey Mondrians. The areas correspond to CoS left (red) and right (green); and IoG left (blue). Talaraich coordinates as given in Table 4.

Table 4: Show a comparison between Cant and Goodale (COS and IOG) Talaraich coordinates and our result (Anterior V4 and Posterior V4 Talaraich coordinates).

	X	Y	Z	Z score	P value
Right V4 (ant)	36	-52.6	-17	4.02	0.000
R CoS (Cant and Goodale)	23	-55	-13		
Left V4 (ant)	-33	-54	-14	5.36	0.000
L CoS (Cant and Goodale)	-37	-42	-15		
Right V4 (post)	33	-80	-10	4.97	0.000
L IOG (Cant and Goodale)	-33	-71	-14		
Left V4 (post)	-32.7	-83	-10	5.17	0.000

Event- related fMRI experiment results

This study is a part of larger scale study. The result shown below is a preliminary result based on the first 6 subjects of that study, but further analysis including these 2 subjects will test its significance.

The event related study is designed to measure the strength of the response of different areas to repeated presentations of the distinct cues. As described above (see Protocol for event related task in Methodology section), we measure the strength through the weight parameter of the fit to the BOLD response. Larger values of the weight in response to one stimulus mean a larger response to that stimulus, or less response suppression. If an area does not show response suppression to repeated presentation of a particular cue, then we conclude it is not selective to that cue. Here we analyse the weights within the areas of interest that were localised previously in the localiser task, namely lateral occipital gyrus, right collateral sulcus and right inferior occipital gyrus.

The result shows that the activation in response to texture change was more than that to changes in colour and shape, especially within LOC and right IOG, while in the right COS the activity was more for texture change and object change than colour changes, and colour change was more in R IOG than LOC (figure 13).

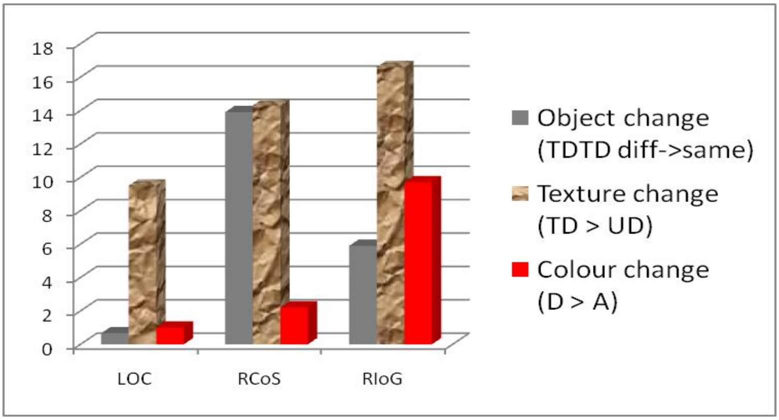


Figure 13: The recorded weight differences in response to the different conditions shown (object, texture and colour changes) within LOC, R COS and R IOG (Group-level analysis of 6 subjects).

Discussion

Colour facilitates natural surface identification

The results from the behavioral experiment indicate that texture alone did not significantly aid in natural object identification compared to other conditions, including color and shape. For instance, accuracy in the texture-only condition was significantly lower across all stimulus onset asynchronies (SOAs) ($p < 0.001$), while the performance with color alone was notably better. However, when color was combined with texture, object identification performance improved significantly.

Additionally, reaction time analysis showed that responses were faster when color and texture were combined compared to either cue alone. Specifically, the average reaction time for the "color + texture" condition was significantly shorter than for texture alone ($p = 0.000$) and color alone ($p = 0.043$). These findings reinforce previous studies suggesting that color facilitates object and scene recognition. For example, Ge et al. (2022) demonstrated that color information plays a crucial role in object classification, emphasizing its importance alongside shape and texture. Similarly, Singh et al. (2020) found that convolutional neural networks (CNNs) rely heavily on color features when making predictions, further highlighting its significance in visual recognition.

Interestingly, the addition of shape to the "color + texture" condition did not significantly enhance performance, suggesting that chromatic texture alone provides sufficient visual information for natural object identification. This aligns with the "surface-plus-edge" theory, which posits that both shape and surface properties contribute to object recognition, in contrast to edge-based theories that prioritize shape over color. Tanaka, Weiskopf, and Williams (2001) illustrated this principle using Monet's painting Water Lilies, which effectively conveys blue water and green vegetation without relying on distinct shape cues.

Diagnostic colour and object identification

It is clear from our behavioural experiment that adding colour to a natural surface significantly facilitates natural object identification, but the question remains whether the colour effect is only related to objects with highly diagnostic colour or does it also include objects without diagnostic colour; for instance, does adding red colour to a car give the same effect as adding red colour to a strawberry? Humphreys et al (1994) suggested that colour facilitates the naming of natural objects, but this effect did not extend to manufactured objects for which colour is far less diagnostic. Many studies suggested that the identification of objects with highly diagnostic colour occurs more than with objects with low diagnostic colour (Tanaka and Presnell 1999).

Another recent study by Singh, Bay, and Mirabile (2020) demonstrated that colour information is used to identify objects, especially those with high diagnostic colour. Tanaka and Presnell (1999) suggested that a "colour-first strategy" is employed when colour information is highly diagnostic, whereas for the identification of low colour diagnostic objects (LCD) a "shape-first strategy" is employed.

Object and set effect were examined to test whether objects with a highly chromatic texture can be identified more effectively than objects with less diagnostic chromatic texture and a carrot was used in every set to serve as control. The performance was significantly better with set 3 (strawberry, Kiwi and carrot) than set 1 (banana, courgette and carrot) over the 3 SOAs and the performance of the carrot in set 3 was significantly better than the performance of the carrot in set 1. We suggested that distinguishing the carrot from the strawberry and kiwi in set 3 was easier than distinguishing the carrot from the courgette in set 1 because they have the same shape outline and less diagnostic chromatic texture compared to the strawberry and kiwi. We suggested from this result that object identification is more effective for objects with high diagnostic chromatic texture (colour and texture together). In addition, the performance was significantly better with the kiwi than the performance with potato in set 2. We suggested from this result that the performance is influenced by the diagnosticity of object attributes.

Information about shape is processed faster than colour and texture

Backward masking was used for timing the visual process at 3 SOAs (40, 100 and 250). An Anova test showed a significant difference between SOA 40 and SOA 100 (p=0.023) and SOA 40 and SOA 250 (p= 0.001), while there were no significant differences between SOA100 and SOA250. A Tukey test showed that the worst performance occurred at SOA 40. We conclude that the mask effect was significant in interrupting the visual information processing at an early stage and in later stages most of the visual information processing has been completed; therefore, the effect of the mask was not significantly different at SOA 100 and SOA 250. Previous studies also recorded that the accuracy and ERP amplitude increased by increasing the delay between a target and a mask (Bacon-Macé, Macé et al. 2005) and also many studies suggested that 250ms is enough to categorise visual stimuli such as faces and animals and produce behavioural action (VanRullen and Thorpe 2001; Rousselet, Macé et al. 2003) and only 150 ms was needed to perform visual processing as recorded with ERP (Thorpe, Fize et al. 1996; Rousselet, Fabre-Thorpe et al. 2002).

The performance for shape alone peaked at SOA 100 which perhaps means that processing of shape information needs less than 100 ms (because it includes behaviour action time). Shape performance was significantly better than the performance for colour and texture alone over SOA 40 and SOA 100, so we can conclude from this that the visual information processing of shape is faster than for texture and colour. Colour performance was significantly increased by increasing mask presentation time and peaked at SOA 250 (this means colour information analysis needs around 250 ms) while the performance for texture alone peaked at SOA 100. This result suggests texture needs a shorter time than colour for analysis, but the performance for colour was significantly greater than for texture at all SOAs. We can suggest that processing of colour is more efficient than texture. A notable study by Cavina-Pratesi et al. (2010), in their recent fMRI studies on form, colour and texture, suggested that the extraction of information about object colour seems to occur relatively early as compared with the extraction of information regarding surface texture. Also, we can understand this result as indicating that colour needs a longer time than texture, but colour performance was significantly better than texture because texture alone did not significantly help in the identification.

Object and surface properties area in the brain

The colour localiser task revealed that significant clusters ($p < 0.000$) were localised within anterior and posterior V4 areas that responded better to coloured versus grey Mondrians. Anterior V4 area is the equivalent of Cant and Goodale's COS and posterior V4 area is the equivalent of Cant and Goodale's IOG. The areas correspond to the Inferior Occipital Gyrus (IOG) and collateral sulcus (COS) areas, identified by Cant and Goodale as being responsible for "surface property" analysis. fMRI studies in humans and monkeys also found an area located posterior in the lingual gyrus (V4) that responds to colour stimulus exchange (Shipp and Zeki 1985; Zeki, Watson et al. 1991; McKeefry and Zeki 1997), but not to objects and faces (Brewer, Liu et al. 2005). As of February 2025, there have been no significant studies published after (2020) that specifically investigate the temporal sequence of object color and texture processing in the human visual cortex. The most recent comprehensive research in this area remains the work by Cavina-Pratesi et al. (2010), which localised two separate areas within the collateral sulcus that were activated separately by colour and texture cues, one located in the medial occipitotemporal cortex (within the anterior collateral sulcus and the lingual gyrus) that was activated mainly by colour, and the other located in the lateral occipitotemporal cortex (within the posterior collateral sulcus) that was activated mainly by texture.

Furthermore, significant activation ($p < 0.000$) was recorded within the right and left lateral occipital area (LOC), or the "object area" as Cant and Goodale described it. The locations of our clusters for the object and colour tasks were very similar to those in Cant and Goodale (2007). Recent neuroimaging studies have reinforced the role of the lateral occipital cortex (LOC) in object recognition. For instance, a study by Liu et al. (2021) demonstrated that the LOC is selectively activated by object shapes, indicating its critical involvement in mid-level visual processing. Additionally, research by Chen et al. (2022) showed that the LOC is interconnected with other face and object processing regions, such as the fusiform face area (FFA) and the occipital face area (OFA), forming a network that facilitates the perception of complex visual stimuli. These findings corroborate earlier research by Cant and Goodale (2007) and Kourtzi and Kanwisher (2000), which identified significant activation within the LOC during object recognition tasks

Conclusion

Our findings indicate that colour plays a more crucial role than texture in natural object identification. While texture alone did not significantly enhance object recognition, combining colour with texture markedly improved performance compared to either cue alone. The addition of shape information to colour and texture combined did not yield significant improvements, suggesting that chromatic texture provides sufficient cues for natural object identification. The FMRI scans showed three areas within the visual cortex to be differentially activated by colour and form: the COS; the IOG (colour localiser); and the LOC (lateral occipital cortex; object localiser). The primary result from the event-related task suggests that activity within LOC, IOG and COS is not selective for single cues, but is modulated by other cues, and that therefore shape, colour and texture interact in object recognition

References

Bacon-Macé, N., M. J. M. Macé, et al. (2005). "The time course of visual processing: Backward masking and natural scene categorisation." Vision Research 45(11): 1459-1469.

Biederman, I. and G. Ju (1988). "Surface versus edge-based determinants of visual recognition." Cognitive Psychology 20(1): 38-64.

Breitmeyer, B. and H. Ogmen (2000). "Recent models and findings in visual backward masking: A comparison, review, and update." Attention, Perception, & Psychophysics 62(8): 1572-1595.

Brewer, A. A., J. Liu, et al. (2005). "Visual field maps and stimulus selectivity in human ventral occipital cortex." Nat Neurosci 8(8): 1102-1109.

Cant, J. S. and M. A. Goodale (2007). "Attention to Form or Surface Properties Modulates Different Regions of Human Occipitotemporal Cortex." Cerebral Cortex 17(3): 713-731.Cavina-Pratesi, C., Kentridge, R.W., Heywood, C.A. & Milner, A.D. (2010) ‘Separate channels for processing form, texture, and color: evidence from fMRI adaptation and visual object agnosia’, Cerebral Cortex, 20(10), pp. 2319–2332.

Ce, L., L. Sharan, et al. (2010). Exploring features in a Bayesian framework for material recognition. Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference.

Chen, X., Wang, Y., & Fang, F. (2022). Functional connectivity of the lateral occipital cortex with face and object processing regions. NeuroImage, 250, 118973.

Ge, Y., Xiao, Y., Xu, Z., Wang, X., & Itti, L. (2022) 'Contributions of Shape, Texture, and Color in Visual Recognition', *arXiv preprint* arXiv:2207.09510.

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

Choi, M., Han, K., Wang, X., Zhang, Y. and Liu, Z. (2023) ‘A dual-stream neural network explains the functional segregation of dorsal and ventral visual pathways in human brains’, arXiv preprint, arXiv:2310.13849

Goffaux, V., C. Jacques, et al. (2004). "Diagnostic Colors Contribute to the Early Stages of Scene Categorization: Behavioral and Neurophysiological Evidence." Journal of Vision 4(8): 873.

Grill-Spector, K. and R. Malach (2004). "THE HUMAN VISUAL CORTEX." Annual Review of Neuroscience 27(1): 649-677.

Hadjikhani, N., A. K. Liu, et al. (1998). "Retinotopy and color sensitivity in human visual cortical area V8." Nature Neuroscience 1(3): 235-241.

Humphrey, G. K., M. A. Goodale, et al. (1994). "The role of surface information in object recognition: studies of a visual form agnosic and normal subjects." Perception 23(12): 1457-1481.

Kandel, E. R. and J. H. Schwartz (1981). Principles of neural science. New York, Elsevier North Holland.

Kourtzi, Z. and N. Kanwisher (2000). "Cortical Regions Involved in Perceiving Object Shape." The Journal of Neuroscience 20(9): 3310-3318.

Liu, P., Zhang, Y., & Xu, Y. (2021). Shape-selective activity in the lateral occipital cortex reflects mid-level visual processing. Journal of Neuroscience, 41(15), 3344–3356.

Malach, R., J. B. Reppas, et al. (1995). "Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex." Proceedings of the National Academy of Sciences 92(18): 8135-8139.

Martinez, A., Valdes-Sosa, M., Javitt, D.C. and Beggel, O. (2024) ‘Disrupted third visual pathway function in schizophrenia: Evidence from real and implied motion processing’, NeuroImage: Clinical, 41, 103570.
McKeefry, D. J. and S. Zeki (1997). "The position and topography of the human colour centre as revealed by functional magnetic resonance imaging." Brain 120(12): 2229-2242.
Mishkin, M., L. G. Ungerleider, et al. (1983). "Object vision and spatial vision: two cortical pathways." Trends in Neurosciences 6: 414-417.
Murray, S. O., H. Boyaci, et al. (2006). "The representation of perceived angular size in human primary visual cortex." Nat Neurosci 9(3): 429-434.
Naor-Raz, G., M. J. Tarr, et al. (2003). "Is color an intrinsic property of object representation?" Perception 32(6): 667-680.
Oliva, A. and P. G. Schyns (2000). "Diagnostic Colors Mediate Scene Recognition." Cognitive Psychology 41(2): 176-210.
Page, W. K., W. M. King, et al. (1994). "Magnocellular or parvocellular lesions in the lateral geniculate nucleus of monkeys cause minor deficits of smooth pursuit eye movements." Vision Research 34(2): 223-239.
Rolls, E. T., M. J. Tovée, et al. (1999). "The Neurophysiology of Backward Visual Masking: Information Analysis." Journal of Cognitive Neuroscience 11(3): 300-311.
Rousselet, G. A., M. Fabre-Thorpe, et al. (2002). "Parallel processing in high-level categorization of natural images." Nature Neuroscience(5): 629-630.
Rousselet, G. A., M. J. M. Macé, et al. (2003). "Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes." Journal of Vision 3(6).
Scabini, L., Sacilotti, A., Zielinski, K. M., Ribas, L. C., De Baets, B., & Bruno, O. M. (2024) 'A Comparative Survey of Vision Transformers for Feature Extraction in Texture Analysis', *arXiv preprint* arXiv:2406.06136.

Shipp, S. and S. Zeki (1985). "Segregation of pathways leading from area V2 to areas V4 and V5 of macaque monkey visual cortex." Nature 315(6017): 322-324.
Tanaka, J. and L. Presnell (1999). "Color diagnosticity in object recognition." Attention, Perception, & Psychophysics 61(6): 1140-1153.
Tanaka, J., D. Weiskopf, et al. (2001). "The role of color in high-level vision." Trends in Cognitive Sciences 5(5): 211-215.
Therriault, D., R. Yaxley, et al. (2009). "The role of color diagnosticity in object recognition and representation." Cognitive Processing 10(4): 335-342.
Thorpe, S., D. Fize, et al. (1996). "Speed of processing in the human visual system." Nature 381(6582): 520-522.
VanRullen, R. and S. J. Thorpe (2001). "The Time Course of Visual Processing: From Early Perception to Decision-Making." Journal of Cognitive Neuroscience 13(4): 454-461.
Wade, A. R., A. A. Brewer, et al. (2002). "Functional measurements of human ventral occipital cortex: retinotopy and colour." Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 357(1424): 963-973.
Webster, M. A. and D. I. A. MacLeod (2011). "Visual adaptation and face perception." Philosophical Transactions of the Royal Society B: Biological Sciences 366(1571): 1702-1725.
Wyszecki, G., V. S. Stiles, et al. (1968). "Color Science: Concepts and Methods, Quantitative Data and Formulas." Physics Today 21(6): 83-84.
Zeki, S. and L. Marini (1998). "Three cortical stages of colour proce
Zeki, S., J. Watson, et al. (1991). "A direct demonstration of functional specialization in human visual cortex." The Journal of Neuroscience 11(3): 641-649.