

Feature-specific interactions in salience from combined feature contrasts: Evidence for a bottom–up saliency map in V1

Department of Psychology, UCL, London, UK,
 Knowledge Creating Communication Research Center,
 NICT, Kyoto, Japan, &
 ATR Computational Neuroscience Laboratories,
 Kyoto, Japan

Ansgar R. Koene

Li Zhaoping

Department of Computer Science, UCL, London, UK



Items that stand out from their surroundings, that is, those that attract attention, are considered to be salient. Saliency is generated by input features in many stimulus dimensions, like motion (M), color (C), orientation (O), and others. We focus on bottom–up saliency generated by *contrast* between the feature properties of an item and its surroundings. We compare the singleton search reaction times (RTs) of items that differ from their surroundings in more than one feature (e.g., C + O, denoted as CO) against the RTs of items that differ from their surroundings in only a single feature (e.g., O or C). The measured RTs for the double-feature singletons are compared against “race model” predictions to evaluate whether saliency in the double-feature conditions is greater than the saliency of either of its feature components. Affirmative answers were found in MO and CO but not in CM. These results are consistent with some V1 neurons being conjunctively selective to MO, others to CO, but almost none to CM. They provide support for the V1 hypothesis of bottom–up saliency (Z. Li, 2002) but are contrary to expectation from the “feature summation” hypothesis, in which different stimulus features are initially analyzed independently and subsequently summed to form a single saliency map (L. Itti & C. Koch, 2001; C. Koch & S. Ullman, 1985; J. M. Wolfe, K. R. Cave, & S. L. Franzel, 1989).

Keywords: saliency, reaction time, feature contrast, feature combinations, saliency map, V1

Citation: Koene, A. R., & Zhaoping, L. (2007). Feature-specific interactions in saliency from combined feature contrasts: Evidence for a bottom–up saliency map in V1. *Journal of Vision*, 7(7):6, 1–14, <http://journalofvision.org/7/7/6/>, doi:10.1167/7.7.6.

Introduction

Items in the visual field with features that are different from their surroundings automatically “pop out” in visual scenes and attract attention. Traditionally, this “pop-out” phenomenon has been demonstrated in singleton search tasks, where the reaction time (RT) for finding targets that differ from uniform surrounding distracters in at least one feature dimension (e.g., color, orientation, and motion) does not increase with the number of distracter items. The degree to which an item or location stands out from its surroundings, that is, attracts attention, is referred to as the “saliency” of the item/location (Titchener, 1908). The term saliency has been used in the visual perception literature in a number of different contexts and with slightly different meanings (e.g., Parkhurst, Law, & Niebur, 2002; Titchener, 1908; Wolfe, 1994). In this article, the term saliency will always refer to a purely bottom–up attraction of attention arising from the *contrast* between the feature properties of an item and its surroundings (Moraglia, 1989; Nothdurft, 1991, 1992, 1993) rather than top–down aspects such as the task-

specific feature relevance. We shall therefore not be considering the effects of top–down modulations that may play a role in guided search (e.g., Bacon & Egeth, 1997; Lamy, Leber, & Egeth, 2004; Leber & Egeth, 2006; Sobel & Cave, 2002; Wolfe, Cave, & Franzel, 1989).

Saliency has been shown to play a crucial role in the localization of targets (Duncan & Humphreys, 1989; Foster & Ward, 1991; Treisman & Gelade, 1980; Treisman & Gormican, 1988; Wolfe, Friedman-Hill, Steward, & O’Connell, 1992), control of eye movements (Deubel & Frank, 1991; Findlay, Brogan, & Wenban-Smith, 1993; Nothdurft & Parlitz, 1993), and allocation of spatial attention (Joseph & Optican, 1996; Julesz, 1981, 1986; Nothdurft, 1999; Wolfe et al., 1989). Furthermore, the global properties of a scene also affect the saliency of a target stimulus; for example, a target item on a nonuniform background is much less salient than a target on a uniform background (Duncan & Humphreys, 1989; Nothdurft, 1991, 1992). Thus, increased overall feature contrast in the background pattern reduces the relative saliency of the target. As shown by various studies (e.g., D’Zmura, 1991; Dick, Ullman, & Sagi, 1987; Duncan & Humphreys, 1989; Itti & Koch, 2001; Koch & Ullman,

1985; Nagy & Sanchez, 1990; Nakayama & Silverman, 1986; Nothdurft, 1993, 1995, 2000; Treisman & Gelade, 1980; Wolfe et al., 1989), salience is generated by feature contrasts in a wide variety of stimulus feature dimensions, like motion, color, luminance, depth, and others. All these saliency effects seem to display qualitatively similar properties as described above. Saliency, therefore, does not appear to be feature specific. This is consistent with the concept that the function of salience is to attract attention for the investigation of the item. The feature properties of the attracting item are investigated after attention has been attracted to that location.

Mechanisms and neural correlates of salience

Various groups have suggested that stimulus information is first processed in separate feature maps, representing single visual features such as red color and vertical orientation, and is subsequently summed into a single master map (Itti & Koch, 2001; Koch & Ullman, 1985; Wolfe et al., 1989) to represent salience irrespective of the actual features (Figure 1a). We shall refer to this hypothesis as the “feature summation hypothesis”. Unfortunately, neither the neural mechanisms nor the exact underlying cortical areas responsible for the feature and saliency maps have been clearly specified in the feature summation hypothesis. The strong influence of the surrounding on a target’s saliency suggests that contextual modulation effects may be particularly important (Nothdurft, 2000). Such contextual effects were demonstrated in a series of studies on single cells in area V1 (e.g., Allman, Miezin, & McGuinness, 1985, 1990; Kastner, Nothdurft, & Pigarev, 1997, 1999; Knierim & van Essen, 1992; Lamme, 1995; Lee, Mumford, Romero, & Lamme, 1998; Nothdurft, Gallant, & van Essen, 1999; Sillito, Grieve, Jones, Cudeiro, & Davis, 1995; Zipser, Lamme, & Schiller, 1996; see also Schofield & Foster, 1995, for a biologically inspired model based on surround suppression). Although the responses to a stimulus in the receptive field (RF) are frequently suppressed when similar stimuli are simultaneously presented outside the RF, the suppression is often weaker or even absent when the surrounding stimuli are different to that in the RF. Thus, the mean responses of the cell population to contrasting stimuli are relatively enhanced over those to uniform texture fields. The response differences correlate well with the salience of pop-out targets. A biologically based model of the preattentive computational mechanisms in the primary visual cortex was developed by Li (2002), showing how V1 neural responses can create a saliency map that awards higher responses to more salient image locations. We shall refer to this hypothesis as the “V1 hypothesis”. Key differences that differentiate the V1 hypothesis (Li, 2002) from the feature summation hypothesis are the following:

1. The V1 hypothesis does not sum the separate feature-based information.

2. The V1 model includes cells that are tuned to specific feature combinations.

The V1 hypothesis relies on conventional V1 cells tuned to input features, such as orientation and color, and known interactions between these cells (Figure 1b).

Model predictions

In the V1 hypothesis, firing rates of V1 cells code salience at the RF locations, regardless of the feature encoded by any specific cell. The activities of multiple V1 cells that respond to the same retinotopic location are not summed. Each V1 cell competes on its own, and the V1 cell with the greatest firing rate determines the most salient location (i.e., winner-take-all). The salience of stimulus features is therefore related to the presence of V1 cells that are sensitive to these features. For example, because a red vertical bar excites “vertical”, “red”, and “red and vertical” sensitive V1 cells at its location, with responses R_O , R_C , and R_{CO} , respectively, from the orientation-tuned (to vertical), color-tuned (to red), and conjunctive-tuned (to red and vertical) cells, the salience of

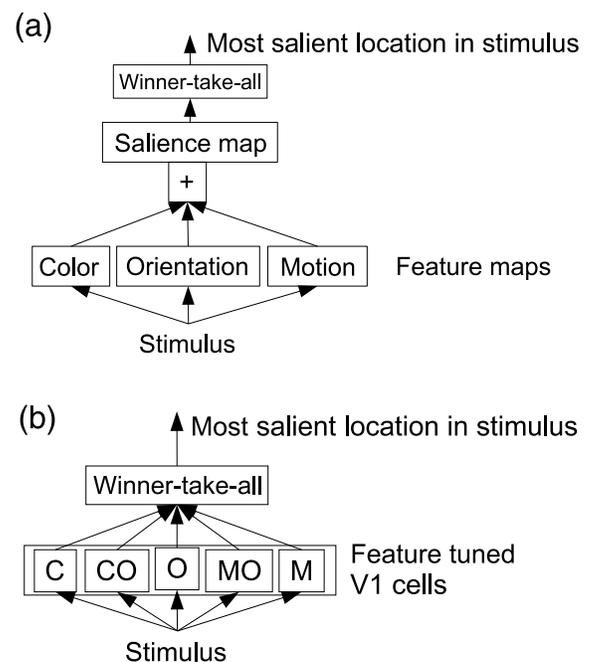


Figure 1. (a) Feature summation hypothesis. Visual inputs are first processed in separate feature maps tuned to different stimulus features (e.g., orientation, color, and motion). The output of these feature maps is summed to produce a single saliency map. (b) V1 hypothesis. V1 cells tuned to different features interact through lateral connections. Activity in cells responding to uniform feature texture stimuli is suppressed through mutual inhibition. The most salient location is the RF location of the cell with the greatest firing rate. C = color, CO = color and orientation, O = orientation, MO = motion direction and orientation, M = motion direction tuned cells.

the red vertical bar is determined by the strongest response from these V1 cells, that is, $\text{Saliency} \propto \max(R_O, R_C, R_{CO})$. When this red vertical bar is among green vertical bars, the red-tuned cell will be the most responsive (because it is least suppressed by other cells tuned to the same color); thus, $\text{Saliency} \propto R_C$. We denote this saliency by unique color as $\text{Saliency}(C)$ and, similarly, saliency by unique orientation as $\text{Saliency}(O)$ and saliency by unique double feature as $\text{Saliency}(CO)$. Then, for a red vertical bar among green vertical bars, $\text{Saliency}(C) \propto R_C$, and similarly, $\text{Saliency}(O) \propto R_O$. However, for a red vertical bar among green horizontal bars, when the singleton differs from the background by both C and O , $\text{Saliency}(CO) \propto \max(R_O, R_C, R_{CO})$ because all three types of cells will escape iso-feature contextual suppression. Consequently, the V1 hypothesis predicts

$$\text{Saliency}(CO) \geq \max[\text{Saliency}(C), \text{Saliency}(O)] \quad (1)$$

This inequality is derived under a simplistic assumption that, to single-feature singletons, the responses from the conjunctively tuned cells are never the dominant one of the responses from all cell types. The inequality still holds when this assumption does not hold, as long as we assume that the conjunctively tuned cells respond more vigorously to the double-feature than the single-feature singleton and that the single-feature tuned cells do not respond more vigorously to the single-feature than the double-feature singletons.

Because in V1, there are numerous conjunctive cells sensitive to “orientation and motion” (MO) or “color and orientation” (CO) but only very few, if any, conjunctive cells sensitive to “color and motion” (CM; Horwitz & Albright, 2005; Hubel & Wiesel, 1959; Livingstone & Hubel, 1984; Ts’o & Gilbert, 1988), the V1 hypothesis makes clearly distinctive predictions for the saliency of MO, as well as CO versus CM combined feature contrasts. If we generalize from above and denote the saliency of a stimulus feature (combination) x as $\text{Saliency}(x)$, where x is color (C), orientation (O), motion (M), or a combination of these (MO, CO, or CM), we can formulate the V1 prediction for the MO and CO feature combinations as

$$\text{Saliency}(MO) \geq \max[\text{Saliency}(M), \text{Saliency}(O)], \quad (2)$$

and

$$\text{Saliency}(CO) \geq \max[\text{Saliency}(C), \text{Saliency}(O)]. \quad (3)$$

For the CM feature combinations, however, the lack of V1 conjunctive cells means that the saliency of a CM feature combination is equal to the saliency of the stronger of the two feature dimensions. Using the same notation as before,

$$\text{Saliency}(CM) = \max[\text{Saliency}(C), \text{Saliency}(M)]. \quad (4)$$

Therefore, comparing the saliency of items that differ from their surrounding by MO, CO, or CM features against the saliency of items that differ from their surroundings in only a single feature, for example, orientation (O), motion (M), or color (C) targets, provides a test of the validity of the V1 hypothesis.

In contrast, the feature summation hypothesis assumes that all stimulus features are initially analyzed in separate feature maps, which are subsequently summed to form a single saliency map. Thus, there is no prior reason to assume that certain feature combinations should be more salient than others. Double-feature items will cause activation in two feature maps, resulting in a greater activation of the summed saliency map at the location of these items. Using the same notation as before, we can write,

$$\begin{aligned} \text{Saliency}(MO) &= w(M) \times \text{Feature}(M) + w(O) \times \text{Feature}(O) \\ &\equiv \text{Saliency}(M) + \text{Saliency}(O) \\ &> \max[\text{Saliency}(M), \text{Saliency}(O)]. \end{aligned} \quad (5)$$

Here, $\text{Feature}(O)$ or $\text{Feature}(M)$ denotes activation in the orientation or motion feature maps, respectively, and $w(O)$ and $w(M)$ are the weights that sum feature maps to the master saliency map (Mueller, Herrer, & Ziegler, 1995). Thus, for example, $\text{Saliency}(O) = w(O) \times \text{Feature}(O)$, and so on, assuming that saliency from a single feature is mainly due to the activation from the corresponding feature map. Therefore, the feature summation hypothesis predicts that

$$\text{Saliency}(MO) > \max[\text{Saliency}(M), \text{Saliency}(O)]. \quad (6)$$

And similarly,

$$\text{Saliency}(CO) > \max[\text{Saliency}(C), \text{Saliency}(O)] \quad (7)$$

and

$$\text{Saliency}(CM) > \max[\text{Saliency}(C), \text{Saliency}(M)]. \quad (8)$$

This model therefore predicts that items that differ from their surroundings in the MO, CO, or CM feature dimensions are all more salient than corresponding items that differ from the surrounding in the C , M , or O feature component alone.

Evidence for increased saliency from feature combinations

Nothdurft (2000) measured the relative saliency of singleton targets defined by various single- or double-

feature contrasts by means of a comparison to reference targets defined by luminance levels. Subjects were briefly (150 ms) presented with two texture arrays of bars, one on either side of the central fixation spot, and had to indicate which texture array contained the more salient target. One array contained a target defined by color, orientation, motion, or luminance feature contrasts (or a combination of these), whereas the other contained the reference target defined by luminance contrast. By varying the luminance contrast of the reference target, Nothdurft measured for each feature (and feature combination) the luminance contrast of the reference that was perceived as equally salient. He thereby provided a feature-independent measure of salience expressed in equivalent luminance contrast levels. From these experiments, Nothdurft concluded the following:

1. Combined feature targets are more salient than single-feature targets.
2. Combined feature targets are less salient than the sum of the component features (suggesting not completely independent processing).
3. The CO feature combination showed the least salience additivity effect.
4. The CM feature combination yielded more salience additivity than the CO feature combination did.
5. The MO feature combination showed about as much salience additivity as the CO feature combination did (perhaps slightly more).

Although Conclusions 3 and 4 seem incompatible with the V1 hypothesis, these conclusions, as well as Conclusion 2, also appear to be incompatible with the feature summation hypothesis.

When comparing Nothdurft's (2000) results against the predictions from the two salience map hypotheses, however, the following caveats must be considered. There are a number of issues with Nothdurft's methodology that may have made the results nonrepresentative for testing the salience models. When attempting to replicate the Nothdurft study, we found that the salience comparison task used in the study proved to be very difficult for many naive subjects, leading to a subject rejection rate of more than 50%. Subjects were rejected if, after careful explanation of the task, their responses proved to be random or completely biased to one feature type (as determined by an inability to fit a psychometric function to their data). In addition, even subjects who could do the task had to be specifically requested to respond quickly to avoid lengthy deliberation within the subject before responding. The task of judging which of two targets was more salient appears to inherently require top-down feature evaluation. Unlike bottom-up salience, the top-down salience or "stimulus priority" that Nothdurft may have measured is strongly affected by stimulus relevance (Bacon & Egeth, 1997; Lamy et al., 2004; Leber & Egeth, 2006; Mevorach, Humphreys, & Shalev, 2006; Sobel &

Cave, 2002). It is therefore not clear whether the task used by Nothdurft provides a reliable measure of the purely bottom-up stimulus salience for which the V1 hypothesis and the feature summation hypothesis give predictions.

The need to find a better method to measure salience was previously raised by Huang and Pashler (2005), who proposed measuring the effect of a distracter (defined by the relevant features) on the RT for finding a target item in a search task. If salience reflects the degree to which an item attracts attention, RT in a search task would seem to offer a more direct measure of item salience. Unfortunately Huang and Pashler did not compare the salience of CM-defined targets versus CO- or MO-defined targets and therefore does not provide us with the desired evidence for testing the proposed salience map hypotheses. In 2002, Krummenacher, Mueller, and Heller measured RT in a search task (oriented bar texture) in which the target could be defined by C, O, or CO. The search RTs revealed a significant reduction for CO-defined targets as opposed to the targets defined by C only or O only. Comparison against race model predictions for independent processing of C and O features revealed that the RTs for the CO feature condition could not simply be accounted for as the results of a winner-take-all race between two independent single-feature processes (similar results were also reported by Zhaoping & May, 2007). Krummenacher et al. concluded that the salience of CO-defined targets is greater than the salience of targets defined by C only or O only. These results are compatible with both the V1 and the feature summation hypotheses. Krummenacher and colleagues are now expanding their work on redundant features to CM and MO as well (private communication, 2007). Unfortunately, a direct comparison with the Nothdurft (2000) study is not possible because his methods of measuring saliency does not provide the possibility for testing against a "race model" while the additivity measure used by Nothdurft is not applicable to the RT data from Krummenacher et al.

None of the published data in the literature provides the crucial information for testing the predictions of the two saliency map hypotheses concerning the salience of feature combinations. We therefore measured RTs in a search task for C-, M-, and O-defined targets and targets defined by any combinations of these features. Similar to Krummenacher et al. (2002), we compare the measured RTs against race model predictions to evaluate if the salience in the double-feature conditions is greater than the salience of either of its feature components.

Hypothesis testing by means of a race model

In simple RT tasks (e.g., Donders, 1868), where participants must respond as quickly as possible to the presentation of any stimulus, responses are faster, on average, when two stimuli are presented than when only

one is presented (e.g., Raab, 1962). One possible cause is statistical facilitation.

If all stimuli (stimulus features) are detected separately, the response is initiated as soon as the first one is detected. The RT is determined by the latency of a single detection process if only one stimulus is presented but is determined by the winner of a race between two detection processes in redundant trials to detect either of the two simultaneously presented stimuli. This is equivalent to the application of the V1 hypothesis to the CM double feature, when the salience is determined by the higher of the two responses to the two single features. Generally, the average time for the winner of the race will be shorter than the average time for either racer (see Miller & Ulrich, 2003).

Alternatively, the activations produced by the two stimuli (stimulus features) could be summed (e.g., the salience map in the feature summation hypothesis) so that the “decision threshold” is reached more rapidly when two stimuli (stimulus features) are presented than when only one is (e.g., Miller, 1982; Schwarz, 1989, 1994; see Townsend & Nozawa, 1995, for an in-depth analysis). One way to distinguish summation/interaction at the stimulus processing level from statistical facilitation is by means of the so-called race model inequality (Miller, 1978, 1982; Ulrich & Giray, 1986), in which

$$F_r(t) \leq F_1(t) + F_2(t), \quad (9)$$

for every value of t , where F_1 and F_2 are the cumulative probability distributions of RT in the two single-stimuli conditions and F_r is the cumulative distribution function (CDF) of RT in the redundant-stimulus condition. This inequality holds for all *separate-activations* race models (e.g., CM in the V1 hypothesis but not in the feature summation hypothesis), where the processes detecting the two possible stimuli (stimulus features) operate separately and each operates at the same speed regardless of whether the other signal is presented (Ashby & Townsend, 1986; Luce, 1986).

Comparison of observer RTs for double-feature stimuli against the corresponding prediction from the race model will therefore determine if there is bottom-up interaction that increases the salience of double-feature-defined targets. We can thus test the basic assumptions of the V1 and feature summation hypotheses independent of any particular computational implementation of these hypotheses. In particular, the feature summation hypothesis predicts that the mean RT for the double features should be shorter than that predicted by the race model regardless of the underlying feature dimensions. In contrast, the V1 hypothesis predicts that mean RTs for CM double-feature stimuli should be predicted by the race model from the RTs for the corresponding single features and that mean RTs for CO and MO double features should be shorter than predicted by the race model.

Methods

Participants

Eight observers (both authors and six naive subjects) participated in the experiment. All had normal or corrected-to-normal vision. The subjects were four women and 4 men. Informed consent was obtained after the nature and possible consequences of the study were explained.

Apparatus and stimuli

The stimuli were presented on a 19-in. Mitsubishi Diamond Pro 2070SB monitor (120 Hz frame rate) and were generated using a Cambridge Research Systems Visage card controlled with Matlab 6.5 (the MathWorks) running on a Pentium 4 PC with Windows XP operating system. Responses were given by means of a USB-numeric keypad using the psychophysics toolbox (Brainard, 1997; Pelli, 1997). The participant sat 57 cm from the stimulus monitor in a quiet, dark room. The visual stimulus consisted of a matrix of 30×22 bar stimuli (Figure 2 shows part of the stimulus screen). The size of each bar in visual angles is approximately 1° long and 0.2° wide. The arrangement of the bars was randomly jittered, giving a horizontal distance between bars varying from 1.2° to 3.3° and a vertical distance varying from 1.1° to 2° visual angle. All bars were colored (green or purple of equal color saturation, in opposite CIE 1976 direction from neutral white on an axis going through $u' = 0.15$, $v' = 0.52$ [green] and $u' = 0.25$, $v' = 0.4$ [purple]) with equal

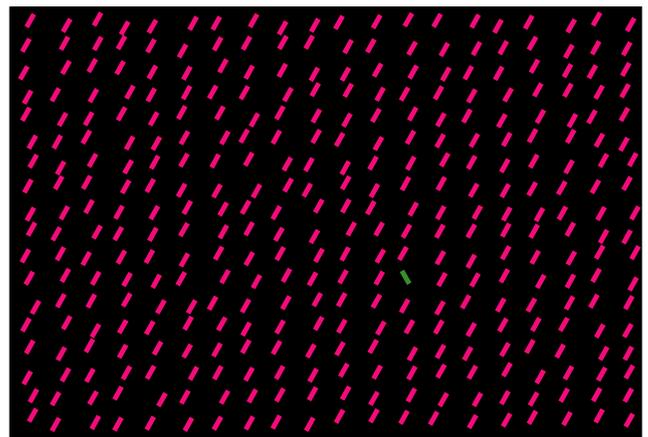


Figure 2. Example of stimulus screen for the color and orientation target feature condition. In addition to the color and orientation features, all bars also uniformly moved horizontally to the left or the right. (Note that, to keep the bar elements clearly visible, only a part of the total stimulus screen is shown here).

luminance (14 cd/m^2), equally tilted clockwise or counter-clockwise from vertical, and moved with the same speed to the left or to the right. On each trial, all of the background distracter bars had the same color, tilt, and motion direction, whereas the target bar had the opposite color, tilt, or motion direction, or a combination of these. Target bars could be at 1 of 18 locations (9 left, 9 right) at a constant eccentricity of 12.8° from the intertrial fixation spot in the center of the screen. Unlike the Nothdurft (2000) and Krummenacher et al. (2002) studies, where the contrast between the target-defining feature property and the corresponding background feature property was always set to the maximum achievable contrast level (i.e., 90° orientation contrast, fully saturated color features, and rapid motion), the degree of color saturation, tilt, and speed of motion was adjusted per individual to achieve single-feature target contrasts that would yield mean search response times of approximately 600 ms (orientation contrast ranged from 20° to 45° , color saturation ranged from 60% to 100%, and motion speed ranged from $0.7^\circ/\text{s}$ to $2.7^\circ/\text{s}$). After each response button press, the bars were replaced by low-contrast crosses and one high-contrast fixation point in the middle of the screen. This fixation screen stayed until the subject pressed another button to initiate the next trial.

Design and procedure

The measurement procedure employed a 2AFC search task in which subjects had to report in which half of the screen (left or right) a singleton target was located. Each trial contained a target that could be defined either by one feature (C, O, or M) or by multiple features (CM, CO, or MO). Targets defined by different feature dimensions were randomly mixed so that subjects could not anticipate which feature dimension to attend to. Each session contained 20 trials for each of the six feature conditions, giving a total of 120 trials per session. Each subject performed about 16 sessions, yielding a total of about 320 trials per feature condition per subject. Subjects performed all sessions consecutively in 1 day. On average, each session lasted 10 min. Subjects were allowed to take short breaks between sessions. Trials were self-paced, and subjects were instructed to take a break between trials if they felt that they were losing concentration. The experimenter sat next to the subjects throughout the experiment to monitor their performance. If a subject seemed to lose concentration (i.e., a sequence of incorrect responses were given), the experimenter suggested taking a short break between trials. Subjects were instructed to respond as rapidly and accurately as possible to indicate which side the target was on. As an extra incentive for giving rapid correct responses, subjects were presented with a score at the end of each

session based on their mean RT and percentage correct responses. Reaction time was recorded as the time between stimulus onset and button press. Subjects were instructed to press the left cursor (located on their left hand side) to indicate “target in left half of screen” and the right cursor (on their right hand side) to indicate “target in right half of screen”. The subject-specific feature contrast tuning was done during separate training sessions before data collection, during which only single-feature-defined targets were used. These sessions also served to familiarize subjects with the search task. We aimed for a mean RT of 600 ms for all single-feature-defined targets to ensure roughly equal task difficulty (saliency) for each single-feature condition, to allow sufficient range for possible significant RT reductions in the double-feature conditions (minimum RT, as measured by presentation of a target stimulus with no background distracters, was approximately 300 ms), and to reduce the available time for the top-down signals to affect the response. A mean RT of 600 ms also reduced the chance for target-distractor collisions in the conditions that contained motion contrast (M, CM, and MO). Depending on the subject, the percentage of trials in which a collision occurred varied from 0% to 2.9% of trials containing motion contrast.

Outlier removal

Outliers were defined as RTs that were more than 3 *SD* different from the mean RT. For each subject and each stimulus condition, the RT outliers were removed based on the mean and standard deviation of the RTs of that particular subject and condition. For each subject and stimulus condition, outliers constituted less than 3.2% of the RT data. All results presented here are based on the RT data after removing the trials with erroneous responses or RT outliers. Qualitatively, the same results were obtained when the outliers are retained.

Results

Percentage error rates

Table 1 lists the percentage error rates per subject per stimulus condition. In all cases, the percentage of incorrect responses never rose above 8%, indicating that subjects were responding with a high degree of accuracy.

Comparison of mean RTs

For each subject, the mean RT was determined for each of the six features/feature combinations. Figure 3 shows

Subject	Percentage of incorrect responses					
	C	O	M	MO	CO	CM
Z.L.	3.2	0.6	5.3	1.2	0.6	5.6
A.K.	3.8	4.7	5.3	1.2	1.9	5.3
C.F.	3.8	5.6	3.4	0.3	1.9	0.3
J.C.	5	6.5	0.6	3.1	2.5	0.3
N.L.	5	1.7	0.6	3.9	1.0	0
R.K.	8	3	3.1	2.8	0.3	2.5
S.A.	4.4	0.3	4.7	1.6	0.7	1.3
S.D.	2.8	4.4	5.9	1.3	3.4	0.6

Table 1. Percentage error rates per target feature condition.

the mean of these mean RTs across subjects. The raw data suggest that subject performance is faster for all double-feature conditions than for any of the single-feature conditions. However, as discussed in the Introduction section, further analysis (by means of a race model) is necessary to determine if this performance difference reflects feature summation or a statistical facilitation in independent parallel processing of stimulus features.

RT cumulative distribution functions

Next, we considered in more detail the RT distributions for each of the stimulus conditions. To get a general comparative overview of the RT distributions, we plotted the CDFs of the RTs by pooling trials from all subjects for

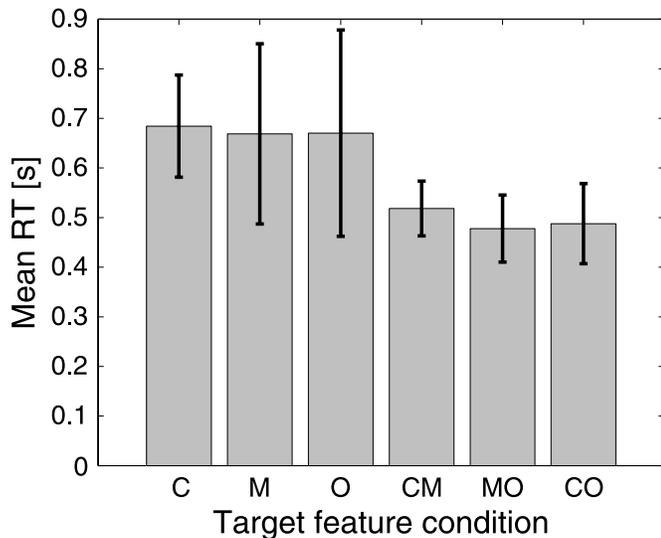


Figure 3. Mean RTs for all eight subjects was averaged, giving the mean and standard deviation between subjects. RTs for the three single-feature and three double-feature target conditions are shown. Error bars show standard deviations.

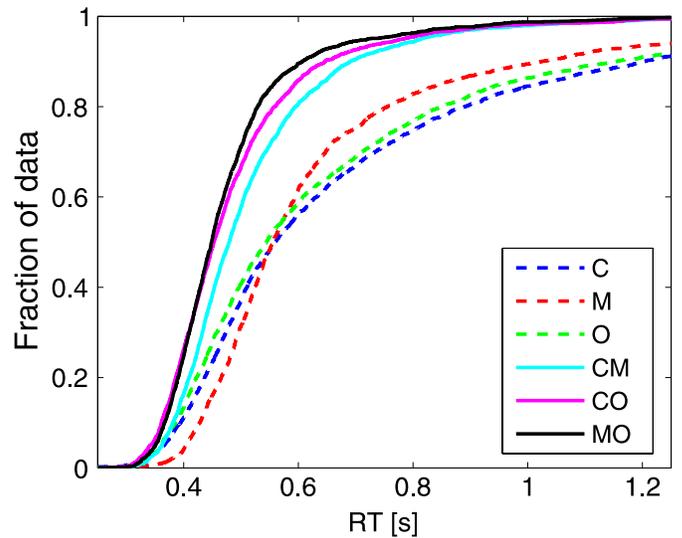


Figure 4. RT CDFs for the three single-feature target conditions (blue, red, and green dashed lines) and the three double-feature conditions (cyan, magenta, and black solid lines). Data were pooled from all eight subjects.

each of the stimulus conditions (Figure 4). The CDFs clearly show a separation between the single- and double-feature conditions, indicating that the double-feature conditions (solid lines) contained a greater percentage of fast RTs, and a separation between the CM condition (cyan line) and the faster MO (black line) and CO (magenta line) conditions.

Comparison against race model prediction

To test whether the decrease in mean RT for the double-feature conditions is indicative of an interactive increase in salience or is the result of independent parallel processing of two features, we compared the RT distributions for the double-feature conditions against the corresponding predictions from a race model. The race model assumes that both features are processed independently and the RT for finding the target corresponds to the time needed for the faster of the two feature processes to reach a decision threshold. To produce race model predictions for the double-feature condition RT distributions, we used a Monte Carlo simulation as follows. In each simulated trial for the double-feature condition (e.g., CM), we randomly selected two RTs, one from each pool of experimental RT data from the constituent single-feature condition (e.g., C and M), and the RT of this double-feature trial is then the shorter of these two selected RTs. To minimize race prediction variance, we simulated 500,000 trials. Figures 5a, 5b, and 5c show the race-model-predicted and real RT distributions for the three double-feature conditions (data pooled

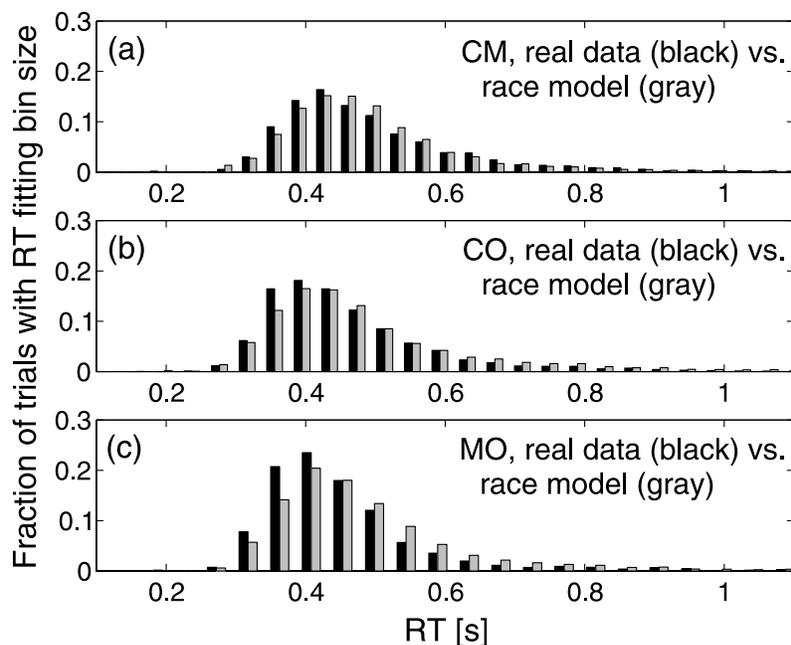


Figure 5. Real RT distributions (pooled data of all eight subjects) and race model simulations of the double-feature target conditions. (a) Color and motion combination. (b) Color and orientation combination. (c) Orientation and motion combination.

from all eight subjects). Figure 6 shows the corresponding CDF plot. The CDF plot clearly shows that the real RT data for the MO and CO feature combinations have a greater percentage of RTs below 600 ms than predicted by the race model. For the CM feature combination, however, this is not the case. Figures 7a and 7b show a comparison between the real and race-model-predicted

mean RTs, that is, $RT_{(real)}$ and $RT_{(race)}$, for the double-feature conditions when the mean RTs for each subject are obtained (shown in Figure 7a) before they are averaged between subjects (Figure 7b). To facilitate comparison, we obtain for each subject

$$\text{mean } RT_{(real)} \pm 95\% \text{ CI } RT_{(real)} - \text{mean } RT_{(race)}, \quad (10)$$

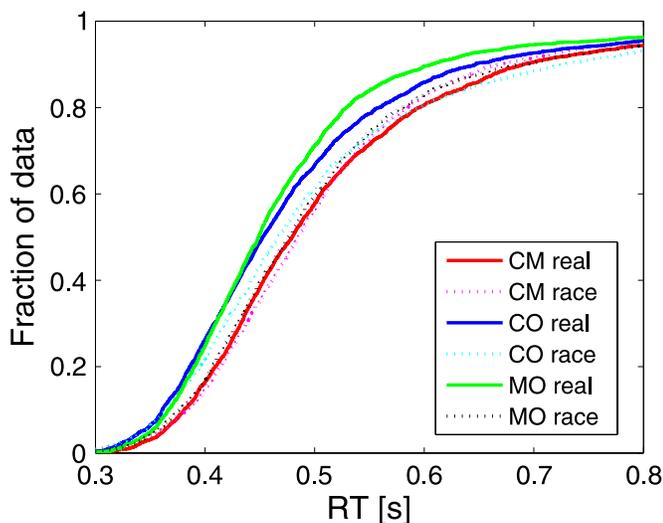


Figure 6. RT CDFs of the subject performance (solid red, green, and blue lines) and race model prediction (dotted magenta, cyan, and gray lines) for the double-feature target conditions (pooled over all eight subjects).

where 95% CI RT denotes 95% confidence interval of the RT and this is plotted in Figure 7a. Figure 7b plots the average value of this quantity above across subjects. Figure 7a shows that for most subjects, mean $RT_{(race)}$ is slower than mean $RT_{(real)}$ for the MO and CO conditions, whereas for the CM condition, mean $RT_{(race)}$ was often faster than mean $RT_{(real)}$. Repeated measures ANOVAs comparing $RT_{(real)}$ against $RT_{(race)}$ for the eight subjects showed no significant difference for CM ($df = 1$, $F = 0.22$, $p = .65$), whereas the CO and MO conditions did show significant differences ($df = 1$, $F = 5.99$, $p = .044$ and $df = 1$, $F = 9.94$, $p = .016$, respectively). Figure 7b shows that, averaged over the eight subjects, the difference between mean $RT_{(race)}$ and mean $RT_{(real)}$, that is, mean $(\text{mean } RT_{(real)} - \text{mean } RT_{(race)})$, was significant for MO (one-sample t test: $t = 3.1532$, $df = 7$, $p = .0161$) and CO (one-sample t test: $t = 2.4469$, $df = 7$, $p = .0443$) but not for CM (one-sample t test: $t = 0.4690$, $df = 7$, $p = .6533$). If we apply a Bonferroni correction, to compensate for the fact that these data might not be fully independent because they were collected in the same sessions, the threshold

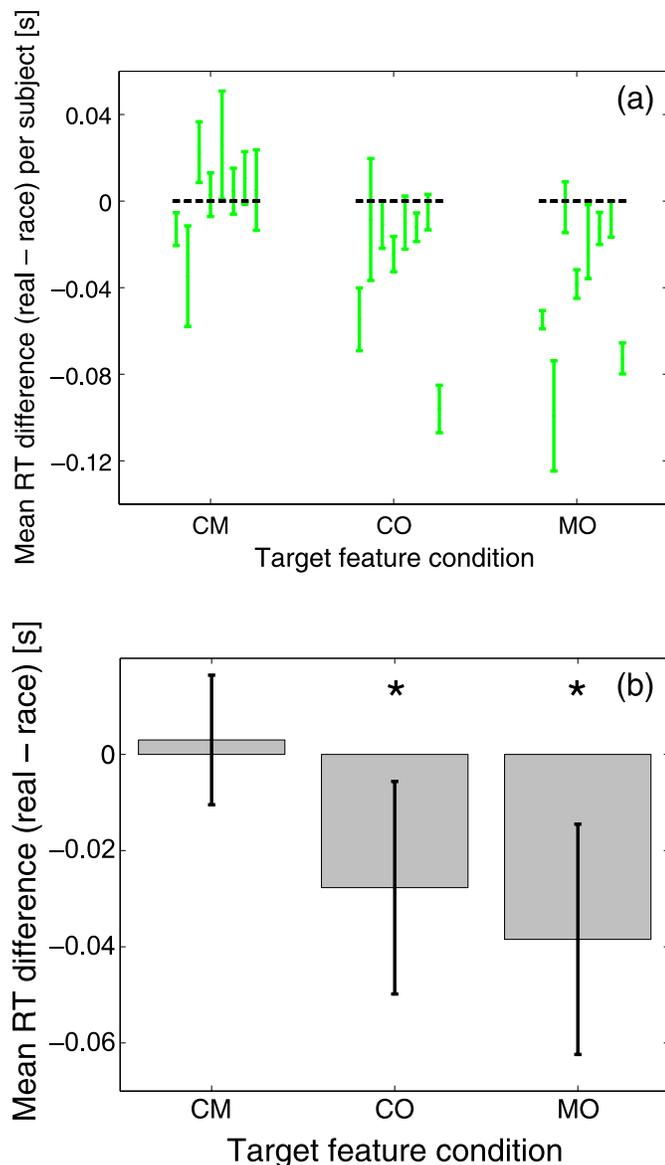


Figure 7. To facilitate comparison across subjects, we show the differences between the mean $RT_{(real)}$ and the race model predicted RT for double-feature singletons. (a) For each of the eight subjects, in green, is mean $RT_{(real)} \pm 95\% \text{ CI } RT_{(real)} - \text{mean } RT_{(race)}$. (b) The difference mean $RT_{(real)} - \text{mean } RT_{(race)}$ averaged across subjects. Error bars show 95% CI. No error bars are given for $RT_{(race)}$ because these can be made arbitrarily small by increasing the number of simulated trials.

values become $p = .0166$, leaving only MO significant. Furthermore, comparison of the difference “mean $RT_{(race)} - \text{mean } RT_{(real)}$ ” for the three double-feature conditions, using matched t test, indicates a significant difference between CM and CO and between CM and MO, but no difference between CO and MO ($p = .03$, $p < .01$, and $p = .34$, respectively). This suggests that O and C, as well as O and M, interact to increase the

salience of targets defined by these feature combinations, whereas C and M features are processed separately, leading to no special boost in target salience.

Discussion

To summarize the findings from this experiment, we have shown by comparison with race model predictions that the RTs to find targets that are defined by CO or OM features are significantly shorter than would be predicted by independent parallel processing of either of the constituent features alone. In contrast, the RTs for CM-defined targets are not significantly different from the predictions from independent parallel processing of the constituent single-feature contrasts.

Comparison with literature

In the CO condition, our finding agrees with that by Krummenacher et al. (2002). They concluded from their CO results that “... there is coactivation of a common mechanism by target signals in different dimensions....”

The results from our study and from the Nothdurft (2000) study, however, diverge on a number of key issues. Whereas we found strong search advantages for MO- and CO-defined targets, but only weak advantages for CM-defined targets, Nothdurft reported that the CO feature combination showed the least salience increase (over that of the single-feature targets), whereas both the MO and the CM feature combinations showed a greater salience increase.

The most probable reason for differences in our data and those of Nothdurft (2000) is the methodological differences of using search task RT measurement versus salience comparison judgments. As discussed in the [Introduction](#) section, when attempting to replicate Nothdurft’s experiment, many of our subjects were not able to successfully perform the task as instructed. Nevertheless, we did qualitatively replicate Nothdurft’s results.

Although the RT method used in our experiment was geared specifically toward measuring bottom-up salience effects by avoiding the need for higher level image processing (i.e., we used singleton pop-out with mean response times of 600 ms or less, and the subjects were typically unaware of the singleton type right after button responses to a trial, which was irrelevant for the task), Nothdurft’s paradigm may inherently require stimulus evaluation at a higher level of processing to compare the salience of the targets presented in the texture arrays in the right half and left half of the screen.

The possible contributing factors toward top-down processing involvement are the following:

- The separation into two distinctly separate stimulus fields. Without this separation, the experiment would

have corresponded more to a uniform “target versus distracter” type task, similar to the task used by Huang and Pashler (2005; except that both singletons were equally valid as target).

- The response method and instruction. Nothdurft’s task would have been more “bottom–up saliency oriented” if the direction of initial eye movements had been recorded while subjects were instructed to simply search for a singleton element. Instructing the subject to “press the button on the side where the more salient target is” promotes responses based on awareness of target features and internal reflection about the percept. The task in Nothdurft’s experiment implies a comparison between two targets, whereas bottom–up saliency is thought to simply provide initial attention grabbing.

Models of bottom–up saliency

The feature summation hypothesis proposes a hierarchical system in which different stimulus features are processed in independent parallel feature maps that are subsequently summed into a master saliency map. Due to this summing stage, it is predicted that all double-feature targets should be more salient than both of the corresponding single-feature-defined targets. To accommodate our data for the CM feature targets, feature summation models would have to introduce the following two fundamental changes: (1) add a layer of computation at which CO and MO features are summed but CM features are not summed, and (2) replace the final summation stage by a winner-take-all stage. Crucially, the CM results show that not all features are summed to form a single saliency map (Figure 8). Furthermore, if Change 2 to the feature summation hypothesis is not included, then summing the outcomes from the CO and MO maps would amount to summing C and M, in contradiction to our data.

In contrast, the V1 hypothesis is able to account for our data without requiring any changes. The RT results

for the different feature conditions in our experiment are consistent with the occurrences of conjunctive cells in V1 that are sensitive to CO or MO feature combinations but not to CM feature combinations (Horwitz & Albright, 2005; Hubel & Wiesel, 1959; Livingstone & Hubel, 1984; Ts’o & Gilbert, 1988), thus validating the predictions made by the physiology-based V1 model of bottom–up saliency (Li, 2002).

By construction, our data do not distinguish the V1 hypothesis from the augmented feature summation hypothesis of bottom–up saliency. Structurally, the augmented feature summation hypothesis has an intermediate stage for CO and MO as subsequent to the single-feature maps. This is so as to retain its resemblance to the original feature summation hypothesis. However, the augmented feature summation hypothesis could be modified such that all the single- and double-feature maps cohabit in a single stage or area, then it would be even structurally indistinguishable from the V1 hypothesis.

The augmented feature summation hypothesis is a fundamental departure from the original feature summation hypothesis. In particular, with the replacement of the final summation stage by the winner-take-all stage in the augmented feature summation hypothesis, there is no longer a need for a separate master saliency map in addition to the feature maps. This is because the attentional selection system could simply ignore (or be blind to) the separation between the single- and double-feature maps, treat all units (or neurons) from these feature maps as if they were from a single neural population, find the most active unit or neuron among them, and direct attention to its RF. Then, as far as this attentional selection is concerned, there is no need to separate the neurons tuned to different features into separate feature maps; all the neurons might as well be residing in a single cortical area such as V1 (Zhaoping, 2005, chap. 93; Zhaoping & Dayan, 2006). In addition, it is known that V2 has neurons tuned to all three types of conjunctions of feature dimensions (Gegenfurtner, Kiper, & Fenstemaker, 1996)—CM, CO, and MO—and thus could not have been responsible for the bottom–up saliency in our task given our data. Our data are thus strong evidences in support of the V1 mechanisms against those in higher cortical areas for being responsible for the bottom–up saliency. However, our data are far from conclusively proving the V1 hypothesis and completely rejecting the original feature summation hypothesis. This is because our experiment only looked at one particular aspect of saliency computation using the most basic set of stimulus variations and feature dimensions. It is not unlikely that more complex visual stimuli and features, such as depth and surface features, could call for additional visual selection mechanisms beyond V1, although they are likely to require longer processing and response latencies (He & Nakayama, 1992) and may be viewed as a different class of bottom–up saliency from those considered in this

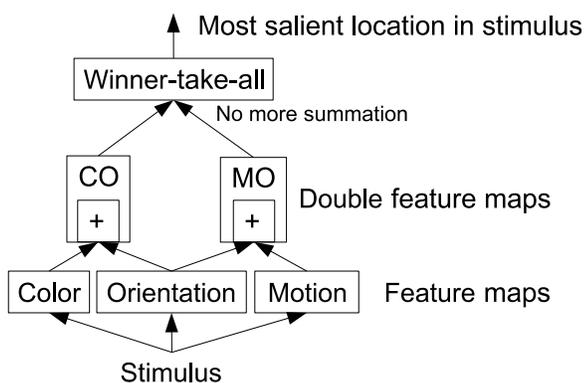


Figure 8. Symbolic sketch of the augmented “feature summation” hypothesis with the required modification to account for our data.

article. With these more complex visual features, some properties in the original feature summation hypothesis could still be relevant. These questions should be closely investigated in future studies.

Role of V1 conjunctive cells for salience

The increased salience of double-feature-defined targets is correlated with the occurrence of conjunctive cells in V1 sensitive to specific feature combinations. If the presence of conjunctive cells makes corresponding double-feature targets more salient, why do targets defined by MO or CO feature combinations not pop out in conjunction search tasks?

The Li (2002) V1 hypothesis is based on the observation of contextual influences on the neural firing rates of V1 cells through finite range lateral interaction between Layer 2–3 pyramidal cells and interneurons. The output activity of each cell therefore depends not only on its direct input (i.e., the stimuli within its classical RF [CRF]) but also on the contextual stimuli outside the CRF. Each cell receives inhibitory inputs from cells with neighboring CRFs that are sensitive to the same feature—iso-feature suppression (with the exception of collinear edge- or bar-detector cells, which provide excitatory input). Although reading out the actual input features from the V1 response requires feature-specific decoding from the population responses (Dayan & Abbott, 2001), the location of the most salient item simply corresponds to the CRF of the most responsive cell. A feature singleton pops out within this framework simply because it evokes response in a neuron that does not suffer from the iso-feature suppression, whereas neurons responding to the nonsingleton items do. To understand why a unique feature conjunction (which is not a feature singleton) does not pop out, V1's conjunctive cells must get inhibitory inputs not only from other conjunctive cells sensitive to the same feature combination but also from single-feature cells sensitive to either of the features to which the conjunctive cell responds. Thus, during conjunctive search tasks, the activity of the conjunctive cell will be suppressed qualitatively (although perhaps not quantitatively) just like the activities in single-feature tuned cells, making the situation qualitatively similar to that without conjunctive cells.

Conclusion

Using RT measurements in a singleton search task, we have shown that OM and CO double-feature-defined targets can be found significantly faster than predicted from a race model assumption of independent processing for the C, M or O features, suggesting a low-level interaction that increases the salience of OM and CO

double-feature-defined targets. For CM double-feature-defined targets, however, the RT was not significantly different from the race model prediction, suggesting that color and motion do not interact (at low level) but are instead processed independently and in parallel. These results are in agreement with predictions from the V1 salience map model (Li, 2002) but contradict the assumption of a general summation over all feature maps that forms the output of the feature maps hypothesis (Itti & Koch, 2001; Koch & Ullman, 1985; Wolfe et al., 1992).

Acknowledgments

This research was supported by a Gatsby Charitable Foundation grant to Dr. Li Zhaoping.

Commercial relationships: none.

Corresponding author: Li Zhaoping.

Email: z.li@ucl.ac.uk.

Address: Department of Computer Science, University College London, Gower St. London WC1E 6BT, UK.

References

- Allman, J., Miezin, F., & McGuinness, E. (1985). Direction- and velocity-specific responses from beyond the classical receptive field in the middle temporal visual area (MT). *Perception, 14*, 105–126. [PubMed]
- Allman, J., Miezin, F., & McGuinness, E. L. (1990). Effects of background motion on the responses of neurons in the first and second cortical visual areas. In G. M. Edelman, W. E. Gall, & M. W. Cowan (Eds.), *Signal and sense: Local and global order in perceptual maps* (pp. 131–142). New York: Wiley-Liss.
- Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review, 93*, 154–179. [PubMed]
- Bacon, W. J., & Egeth, H. E. (1997). Goal-directed guidance of attention: Evidence from conjunctive visual search. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 948–961. [PubMed]
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433–436. [PubMed]
- D'Zmura, M. (1991). Color in visual search. *Vision Research, 31*, 951–966. [PubMed]
- Dayan, P., & Abbott, L. (2001). *Theoretical neuroscience*. Cambridge, MA: MIT Press.
- Deubel, H., & Frank, H. (1991). The latency of saccadic eye movements to texture-defined stimuli. In R. Schmid & D. Zambardi (Eds.), *Oculomotor*

- control and cognitive processes* (pp. 369–384). North Holland: Elsevier.
- Dick, M., Ullman, S., & Sagi, D. (1987). Parallel and serial processes in motion detection. *Science*, *237*, 400–402. [[PubMed](#)]
- Donders, F. C. (1868). Over de snelheid van psychische processen [On the speed of mental processes]. In W. G. Koster (Ed. and Trans.), *Attention and performance II* (pp. 412–431). Amsterdam: North Holland. (Original work published 1868.)
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*, 433–458. [[PubMed](#)]
- Findlay, J. M., Brogan, D., & Wenban-Smith, M. G. (1993). The spatial signal for saccadic eye movements emphasizes visual boundaries. *Perception & Psychophysics*, *53*, 633–641. [[PubMed](#)]
- Foster, D. H., & Ward, P. A. (1991). Asymmetries in oriented-line detection indicate two orthogonal filters in early vision. *Proceedings of the Royal Society B: Biological Sciences*, *243*, 75–81. [[PubMed](#)]
- Gegenfurtner, K. R., Kiper, D. C., & Fenstemaker, S. B. (1996). Processing of color, form, and motion in macaque area V2. *Visual Neuroscience*, *13*, 161–172. [[PubMed](#)]
- He, Z. J., & Nakayama, K. (1992). Surfaces versus features in visual search. *Nature*, *359*, 231–233. [[PubMed](#)]
- Horowitz, G. D., & Albright, T. D. (2005). Paucity of chromatic linear motion detectors in macaque V1. *Journal of Vision*, *5*(6):4, 525–533, <http://journalofvision.org/5/6/4/>, doi:10.1167/5.6.4. [[PubMed](#)] [[Article](#)]
- Huang, L., & Pashler, H. (2005). Quantifying object salience by equating distractor effects. *Vision Research*, *45*, 1909–1920. [[PubMed](#)]
- Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurons in the cat's visual cortex. *The Journal of Physiology*, *148*, 574–591. [[PubMed](#)] [[Article](#)]
- Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2*, 194–203. [[PubMed](#)]
- Joseph, J. S., & Optican, L. M. (1996). Involuntary attention shifts due to orientation differences. *Perception & Psychophysics*, *58*, 651–665. [[PubMed](#)]
- Julesz, B. (1981). Textons, the elements of texture perception, and their interactions. *Nature*, *290*, 91–97. [[PubMed](#)]
- Julesz, B. (1986). Texton gradients: The texton theory revisited. *Biological Cybernetics*, *54*, 245–251. [[PubMed](#)]
- Kastner, S., Nothdurft, H. C., & Pigarev, I. N. (1997). Neuronal correlates of pop-out in cat striate cortex. *Vision Research*, *37*, 371–376. [[PubMed](#)]
- Kastner, S., Nothdurft, H. C., & Pigarev, I. N. (1999). Neuronal responses to orientation and motion contrast in cat striate cortex. *Visual Neuroscience*, *16*, 587–600. [[PubMed](#)]
- Knierim, J. J., & van Essen, D. C. (1992). Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *Journal of Neurophysiology*, *67*, 961–980. [[PubMed](#)]
- Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, *4*, 219–227. [[PubMed](#)]
- Krummenacher, J., Mueller, H. J., & Heller, D. (2002). Visual search for dimensionally redundant pop-out targets: Parallel-coactive processing of dimensions is location specific. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 1303–1322. [[PubMed](#)]
- Lamme, V. A. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *Journal of Neuroscience*, *15*, 1605–1615. [[PubMed](#)] [[Article](#)]
- Lamy, D., Leber, A., & Egeth, H. E. (2004). Effects of task relevance and stimulus-driven salience in feature-search mode. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 1019–1031. [[PubMed](#)]
- Leber, A. B., & Egeth, H. E. (2006). It's under control: Top-down search strategies can override attentional capture. *Psychonomic Bulletin & Review*, *13*, 132–138. [[PubMed](#)]
- Lee, T. S., Mumford, D., Romero, R., & Lamme, V. A. (1998). The role of the primary visual cortex in higher level vision. *Vision Research*, *38*, 2429–2454. [[PubMed](#)]
- Li, Z. (2002). A saliency map in primary visual cortex. *Trends in Cognitive Sciences*, *6*, 9–16. [[PubMed](#)]
- Livingstone, M. S., & Hubel, D. H. (1984). Anatomy and physiology of a color system in the primate visual cortex. *Journal of Neuroscience*, *4*, 309–356. [[PubMed](#)] [[Article](#)]
- Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. Oxford: Oxford University Press.
- Mevorach, C., Humphreys, G. W., & Shalev, L. (2006). Opposite biases in salience-based selection for left and right posterior parietal cortex. *Nature Neuroscience*, *9*, 740–742. [[PubMed](#)]
- Miller, J. O. (1978). Multidimensional same-different judgments: Evidence against independent comparisons of dimensions. *Journal of Experimental*

- Psychology: Human Perception and Performance*, 4, 411–422. [PubMed]
- Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, 14, 247–279. [PubMed]
- Miller, J., & Ulrich, R. (2003). Simple reaction time and statistical facilitation: A parallel gains model. *Cognitive Psychology*, 46, 101–151. [PubMed]
- Moraglia, G. (1989). Display organization and the detection of horizontal line segments. *Perception & Psychophysics*, 45, 265–272. [PubMed]
- Mueller, H. J., Herrer, D., & Ziegler, J. (1995). Visual search for singleton feature targets within and across feature dimensions. *Perception & Psychophysics*, 57, 1–17. [PubMed]
- Nagy, A. L., & Sanchez, R. R. (1990). Critical color differences determined with a visual search task. *Journal of the Optical Society of America A, Optics and image science*, 7, 1209–1217. [PubMed]
- Nakayama, K., & Silverman, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature*, 320, 264–265. [PubMed]
- Nothdurft, H. C. (1991). Texture segmentation and popout from orientation contrast. *Vision Research*, 31, 1073–1078. [PubMed]
- Nothdurft, H. C. (1992). Feature analysis and the role of similarity in pre-attentive vision. *Perception & Psychophysics*, 52, 335–375. [PubMed]
- Nothdurft, H. C. (1993). The role of features in preattentive vision: Comparison of orientation, motion and color cues. *Vision Research*, 33, 1937–1958. [PubMed]
- Nothdurft, H. C. (1995). Generalized feature contrast in preattentive vision. *Perception*, 24, S22.
- Nothdurft, H. C. (1999). Focal attention in visual search. *Vision Research*, 39, 2305–2310. [PubMed]
- Nothdurft, H. C. (2000). Salience from feature contrast: Additivity across dimensions. *Vision Research*, 40, 1183–1202. [PubMed]
- Nothdurft, H. C., Gallant, J. L., & van Essen, D. C. (1999). Response modulation by texture surround in primate area V1: Correlates of “popout” under anesthesia. *Visual Neuroscience*, 16, 15–34. [PubMed]
- Nothdurft, H. C., & Parlitz, D. (1993). Absence of express saccades to texture or motion defined targets. *Vision Research*, 33, 1367–1383. [PubMed]
- Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, 42, 107–123. [PubMed]
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442. [PubMed]
- Raab, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*, 24, 574–590. [PubMed]
- Schofield, A. J., & Foster, D. H. (1995). Artificial neural networks simulating visual texture segmentation and target detection in line-element images. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences*, 350, 401–412. [PubMed]
- Schwarz, W. (1989). A new model to explain the redundant-signals effect. *Perception & Psychophysics*, 46, 498–500. [PubMed]
- Schwarz, W. (1994). Diffusion, superposition, and the redundant-targets effect. *Journal of Mathematical Psychology*, 38, 504–520.
- Sillito, A. M., Grieve, K. L., Jones, H. E., Cudeiro, J., & Davis, J. (1995). Visual cortical mechanisms detecting focal orientation discontinuities. *Nature*, 378, 492–496. [PubMed]
- Sobel, K. V., & Cave, K. R. (2002). Roles of salience and strategy in conjunctive search. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 1055–1070. [PubMed]
- Titchener, E. B. (1908). *Lectures on the elementary psychology of feeling and attention*. New York: The MacMillan Company.
- Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial and coactive theories. *Journal of Mathematical Psychology*, 39, 321–359.
- Treisman, A. M., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, 12, 97–138. [PubMed]
- Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, 95, 15–48. [PubMed]
- Ts’o, D. Y., & Gilbert, C. D. (1988). The organization of chromatic and spatial interactions in the primate striate cortex. *Journal of Neuroscience*, 8, 1712–1727. [PubMed] [Article]
- Ulrich, R., & Giray, M. (1986). Separate-activation models with variable base times: Testability and checking of cross-channel dependency. *Perception & Psychophysics*, 39, 248–254. [PubMed]
- Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin and Review*, 1, 202–238.
- Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model of visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 419–432. [PubMed]

- Wolfe, J. M., Friedman-Hill, S. R., Stewart, M. I., & O'Connell, K. M. (1992). The role of categorization in visual search for orientation. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 34–49. [[PubMed](#)]
- Zhaoping, L. (2005). The primary visual cortex creates a bottom up saliency map. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *Neurobiology of attention* (pp. 570–575). San Diego, CA: Elsevier.
- Zhaoping, L., & Dayan, P. (2006). Pre-attentive visual selection. *Neural Networks*, *19*, 1437–1439. [[PubMed](#)]
- Zhaoping, L., & May, K. A. (2007). Psychophysical tests of the hypothesis of a bottom-up saliency map in primary visual cortex. *PLoS Computational Biology*, *3*, e62. [[PubMed](#)] [[Article](#)]
- Zipser, K., Lamme, V. A., & Schiller, P. H. (1996). Contextual modulation in primary visual cortex. *Journal of Neuroscience*, *16*, 7376–7389. [[PubMed](#)] [[Article](#)]