Affective and contextual values modulate spatial frequency use in object recognition

Laurent Caplette[1], Gregory West[1], Marie Gomot[2], Frédéric Gosselin[1]* and Bruno Wicker[3]

[1] Département de Psychologie, CERNEC, Université de Montréal, Montréal, QC, Canada
[2] INSERM U930 Imagerie et Cerveau, Université François-Rabelais de Tours, CHRU de Tours, Tours, France
[3] CNRS UMR 7289, Institut de Neurosciences de la Timone, Aix-Marseille Université, Marseille, France

Visual object recognition is of fundamental importance in our everyday interaction with the environment. Recent models of visual perception emphasize the role of top-down predictions facilitating object recognition via initial guesses that limit the number of object representations that need to be considered. Several results suggest that this rapid and efficient object processing relies on the early extraction and processing of low spatial frequencies (LSF). The present study aimed to investigate the SF content of visual object representations and its modulation by contextual and affective values of the perceived object during a picture-name verification task. Stimuli consisted of pictures of objects equalized in SF content and categorized as having low or high affective and contextual values. To access the SF content of stored visual representations of objects, SFs of each image were then randomly sampled on a trial-by-trial basis. Results reveal that intermediate SFs between 14 and 24 cycles per object (2.3–4 cycles per degree) are correlated with fast and accurate identification for all categories of objects. Moreover, there was a significant interaction between affective and contextual values over the SFs correlating with fast recognition. These results suggest that affective and contextual values of a visual object modulate the SF content of its internal representation, thus highlighting the flexibility of the visual recognition system.

Keywords: object recognition, internal representations, affective value, context, spatial frequencies

INTRODUCTION

Rapid and accurate visual recognition of everyday objects encountered in different orientations, seen under various illumination conditions, and partially occluded by other objects in a visually cluttered environment is necessary for our survival. The first theoretical efforts to explain this feat relied on purely bottom-up mechanisms in the visual system: cells in early visual areas would be sensitive to low-level features and cells in higher areas would integrate this information in order to then match it to a representation in memory (e.g., Maunsell and Newsome, 1987). However, it is improbable that feedforward pathways alone can account for object recognition because of their severely limited information processing capabilities (Gilbert and Sigman, 2007). Moreover, since these early theoretical efforts, the essential role of such feedback mechanisms in vision has been amply demonstrated (e.g., Rao and Ballard, 1999; Tomita et al., 1999; Barceló et al., 2000; Pascual-Leone and Walsh, 2001). Nowadays, most top-down models of object recognition (e.g., Grossberg, 1980; Ullman, 1995; Friston, 2003) propose that the search for correspondence between the input pattern and the stored representations is a bidirectional process where the input activates bottom-up as well as top-down streams that simultaneously explore many alternatives; object recognition is achieved when the counter streams meet and a match is found. The content of these stored representations could depend on several factors such as task requirements (e.g., perception or action, basic-level

vs. superordinate-level categorization) or categorical properties of the object (e.g., animate vs. inanimate, affective vs. non-affective, social vs. non-social; Logothetis and Sheinberg, 1996). Understanding the properties of the stored representations that lead to the generation of predictions thus is an important unexplored issue. In particular, it remains to be understood if different representational systems are used during recognition of different categories of visual objects.

Building on the predictive account of visual object recognition, Bar (2003) proposed a brain mechanism for the cortical activation of top-down processing during object recognition, where low spatial frequencies (LSFs) of the image input are projected rapidly and directly through quick feedforward connections, from early visual areas into the dorsal visual stream. Such LSF information activates a relatively small set of probable candidate interpretations of the visual input in higher prefrontal integrative centers. These initial guesses are then back-projected along the reverse hierarchy to guide further processing and gradually encompass high spatial frequencies (HSFs) available at lower cortical visual areas. This proposal is supported by neurophysiological, computational and psychophysical evidence that LSFs are processed earlier than HSFs (Watt, 1987; Schyns and Oliva, 1994; Bredfeldt and Ringach, 2002; Mermillod et al., 2005; Musel et al., 2012; for reviews, see Bullier, 2001; Bar, 2003; Hegdé, 2008) and that top-down processing in visual recognition relies on LSFs (Bar et al., 2006); moreover, magnocellular projections, which are

more sensitive to LSFs (Derrington and Lennie, 1984), seem to be implicated in initiation of top-down processing (Kveraga et al., 2007). Stored internal representations may thus be biased toward LSFs, since objects would be primarily matched in memory with an LSF draft.

Only a handful of studies have focused on the effect of specific SF band filtering during object recognition. In a name-picture verification task, low-pass filtering selectively impaired subordinate-level category verification (e.g., verify the "Siamese" category instead of the "animal" category at the superordinate level or the "cat" category at the basic level), while having little to no effect on basic-level category verification, suggesting that basic-level categorization does not particularly rely on LSFs (Collin and McMullen, 2005). On the other hand, Harel and Bentin (2009) reported that subordinate-level categorization was impaired by the removal of HSFs, but also that basic-level categorization was equally impaired by removal of either HSFs or LSFs, thus suggesting that neither of these bands is especially useful for recognition at the basic level. Finally, using a superordinate-level categorization task, Calderone et al. (2013) reported no difference in accuracy or response times between LSFs and HSFs. Overall, these studies suggest that, although this seems a bit different for subordinate-level categorization, neither LSFs or HSFs have a privileged role in object recognition. Even if LSFs do initiate a top-down processing, this suggests that their overall role in recognition is negligible; other SFs (neither low or high), however, may have a preponderant role.

Intrinsic properties of visual objects such as their affective value or contextual associativity may modulate the content of internal representations. Because of their great adaptive value, emotional objects might necessitate fast recognition, to facilitate an immediate behavioral response; this is likely to apply to both dangerous and pleasant stimuli, the former threatening survival and the latter promoting it (Bradley, 2009). In fact, the brain's prediction about the identity of a visual object may be partly based on its affective value, i.e., prior experiences of how perception of a given object has influenced internal body sensations. As such, affective value could be not just a label or judgment applied to the object post-recognition, but rather an integral component of mental object representations (Lebrecht et al., 2012) and could act as an additional clue to the object's identity to facilitate its recognition (Barrett and Bar, 2009). Since emotional objects need to be processed quickly, it is likely that LSFs, which are extracted rapidly, are particularly important for their recognition. In agreement with this idea, there is some evidence that LSFs are more present in representations of objects with strong affective value than in representations of neutral objects. Mermillod et al. (2010) reported that threatening stimuli were recognized faster and more accurately than neutral ones with LSFs but not with HSFs. Other behavioral and neuroimaging studies also suggested an interaction between emotional content and LSFs in various perceptual tasks. Bocanegra and Zeelenberg (2009), for instance, observed that in a Gabor orientation discrimination task, briefly presented fearful faces improved subjects' performance with LSF gratings while impairing it with HSF gratings. Moreover, early ERP amplitudes sensitive to affective content were found to be greater when unpleasant scenes were presented intact or in LSFs rather than

in HSFs (Alorda et al., 2007). In the same vein, Vuilleumier et al. (2003) observed that the amygdala responded to fearful faces only if LSFs were present in the stimulus. In an intracranial ERP study where subjects were presented with both visible and invisible (masked) faces, Willenbockel et al. (2012) found that amygdala activation correlated mostly with SFs around 2 and 6 cycles/face, while insula activation correlated mostly with slightly higher SFs near 9 cycles/face. All these results suggest that the internal representations of objects with affective value would comprise more LSFs than representations of neutral objects.

Relatedly, the contextual associativity of a visual object— "what other objects or context might go with this object?" (Bar, 2004; Fenske et al., 2006)—could also impact on the SF content of its mental representation. It has been shown that recognition of an object that is highly associated with a certain context facilitates the recognition of other objects that share the same context (e.g., Bar and Ullman, 1996). A lifetime of visual experience would lead to contextual associations that guide expectations and aid subsequent recognition of associated visual objects through rapid sensitization of their internal representations (Biederman, 1972, 1981; Palmer, 1975; Biederman et al., 1982; Bar and Ullman, 1996). This associative processing is quickly triggered merely by looking at an object and would be critical for visual recognition and prediction (Bar and Aminoff, 2003; Aminoff et al., 2007). It has been suggested that the rapidly extracted LSFs of an object image are sufficient to activate these associated representations, and thus that the representations of contextual objects are likely to be biased toward LSFs (Bar, 2004; Fenske et al., 2006). However, this hypothesis has never been tested directly.

Affective and contextual values may also interact, so that representations of visual objects with affective value could be modulated by their contextual value or vice-versa (e.g., Storbeck and Clore, 2005; Brunyé et al., 2013; Shenhav et al., 2013). Indeed, the affective value of a given object is often defined by the context to which it has been associated to in memory. For example, a tomb elicits sadness, not because it is inherently sad, but because it evokes a context of cemetery/death. As such, affective objects might be differentially represented whether or not their affective value originates from their associated contexts. Interactions between both psychological properties have been reported. For instance, our affective state influences the breadth of the associations we make (Storbeck and Clore, 2005) and conversely, the generation of associations influences our affective state (Brunyé et al., 2013). Also, it seems that associative and affective processing both take place in the medial orbitofrontal cortex, and that both contextual and affective values might in fact relate to a more unified purpose (Shenhav et al., 2013).

The current study examined the SF content of stored internal representations of visual objects with different affective and contextual values, by evaluating what are the SFs in the stimuli that correlate with fast and accurate identification. Stimuli consisted of pictures of objects equalized in SF content and categorized as having low or high affective and contextual values. The SFs of these stimuli were randomly sampled on a trial-by-trial basis while subjects categorized the objects portrayed in the images. By varying affective value, contextual value and spatial frequencies available in the object image altogether, we aimed to clarify their roles in

visual recognition, and to study potential interactions between them.

METHODS

PARTICIPANTS

Forty-seven healthy participants (33 males) with normal or corrected-to-normal visual acuity were recruited on the campus of the Université de Montréal for an object recognition study. Participants were aged between 19 and 31 years ($M = 23.04$; $SD = 3.13$) and did not suffer from any reading disability. A written informed consent was obtained prior to the experiment, and a monetary compensation was provided upon its completion.

APPARATUS

The experimental program was run on a Mac Pro computer in the Matlab (Mathworks Inc.) environment, using functions from the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). A refresh rate of 120 Hz and a resolution of 1920×1080 pixels were set on the Asus VG278H monitor used for stimuli presentation. The relationship between RGB values and luminance levels was linearized. Luminance depth was 8 bits, and minimum and maximum luminance values were 1.1 cd/m^2 and 134.0 cd/m^2, respectively. A chin rest was used to maintain viewing distance at 76 cm.

STIMULI

Selection and validation

One hundred fifty six object images were pre-selected mainly from the database used in Shenhav et al. (2013) but also from Internet searches. Each object image was presented to 30 raters who decided either (i) if they associated the object to a particular emotion, and if so, to which one or (ii) if they associated the object to a particular context, and if so, to which one. For the experiment, we selected 18 objects with clear consensus (or absence of) regarding their contextual and affective values in each of our four object categories: contextual emotional, non-contextual emotional, contextual neutral and non-contextual neutral (**Figure 1**, Table S1). Clear consensus about high affective or high contextual value meant that an object was associated to the same context or to the same emotion by more than 75% of raters; and clear consensus about low affective or contextual value meant that an object was associated to no particular context or emotion by more than 75% of the raters. Fifty-one of the selected images came from the Shenhav et al. (2013) database, and our affective and contextual ratings for these images closely matched theirs.

Control of low-level features

Stimuli thus consisted of 72 grayscaled object images of 256×256 pixels presented on a mid-gray background. The images subtended $6 \times 6°$ of visual angle. Median object width was equal to 237 pixels. To target our investigation on stored internal representations and get rid of a potential interaction between the visual input and the representation, spatial frequency content and luminance were equalized across stimuli using the SHINE toolbox (Willenbockel et al., 2010a). Resulting images had a RMS contrast of 0.075. We reduced the undesired impact of psycho-linguistic
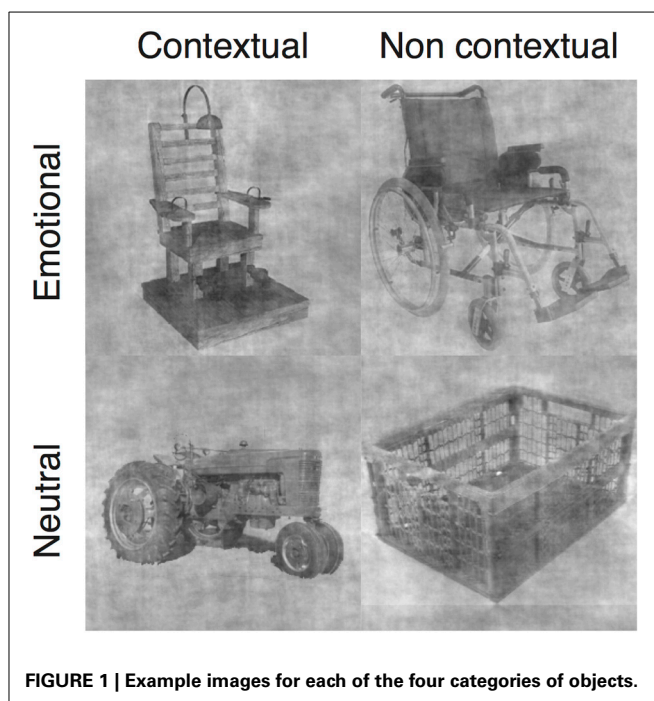

FIGURE 1 | Example images for each of the four categories of objects.
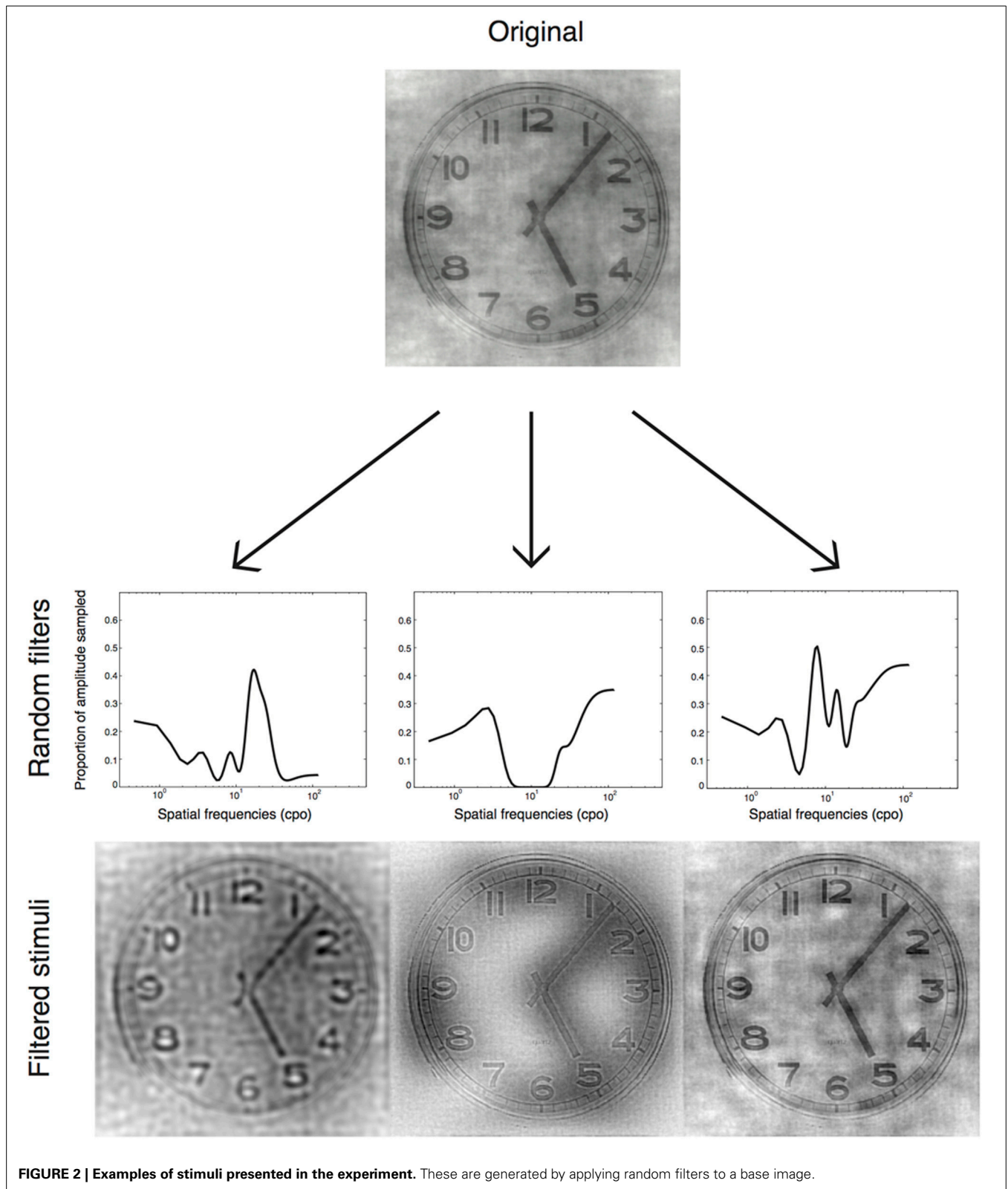
factors, such as word length and lexical frequency, on response times by transforming these into z-scores for every object. For example, we computed the mean and standard deviation of the RTs of the correct positive trials in which the electric chair was presented, and we used these statistics to transform those RTs into z-scores. We did the same for all the other objects. As a result, the means and standard deviations of the RTs associated with every word were strictly identical, and all RT variations due to differences between the words were eliminated.

Sampling

SF content of the images properly padded was extracted via Fast Fourier Transform (FFT) and randomly filtered at each trial, according to the SF Bubbles method (Willenbockel et al., 2010b). In short, each spatial frequency filter was created by first generating a random vector of 10,240 elements consisting of 20 ones (the number of bubbles) among zeros. Second, the resulting vector was convolved with a Gaussian kernel that had a standard deviation of 1.8. Third, the vector was log transformed so that the SF sampling approximately fit the SF sensitivity of the human visual system (see De Valois and De Valois, 1990). The resulting sampling vector contained 256 elements representing each spatial frequency from 0.5 to 128 cycles per image. To create the two-dimensional spatial frequency filtered images, vectors were rotated about their origins and dot-multiplied with the FFT amplitudes (see Willenbockel et al., 2010b, for methodological details). Thus, several SF bandwidths were revealed in each stimulus; and objects were presented several times with different SF bandwidths revealed every time (**Figure 2**).

PROCEDURE

After they had completed a short questionnaire for general information (age, sex, education, language, etc.), participants sat

FIGURE 2 | Examples of stimuli presented in the experiment. These are generated by applying random filters to a base image.

comfortably in front of a computer monitor, in a dim-lighted room. Participants did two 500-trial blocks, with a short break in between. Each trial began with a central fixation cross lasting 300 ms, followed by a blank screen for 100 ms, the SF-filtered random object image for 300 ms, a central fixation cross for 300 ms, a blank screen for 100 ms, and finally a matching or mismatching object name that remained on the screen until the participant had answered or for a maximum of 1000 ms. Subjects

were asked to indicate with a keyboard key press as accurately and rapidly as possible whether or not the name matched the object depicted in the image. This picture-name verification task was chosen because it imposes a specific level of categorization to subjects (we chose the basic-level) without focusing attention explicitly on either affective or contextual value of the object. Name and object matched on half the trials.

SPATIAL FREQUENCY DATA ANALYSIS

To determine the spatial frequencies that contributed most to fast object recognition for each condition, we performed least-square multiple linear regressions between RTs and corresponding sampling vectors. Only correct positive trials (i.e., when the name matched the object, and the participant answered correctly) were included in the analysis. RTs were first z-scored for every object to minimize undesired sources of variability pertaining to psycholinguistic factors such as word length and lexical frequency (see Stimuli: Control of low-level features). They were further z-scored for each condition in each subject's session to diminish variability due to task learning. Trials associated with z-scores over 3 or below 3 were discarded (<1.8% of trials).

We call the resulting vectors of regression coefficients classification vectors. We first contrasted the classification vector for all objects against zero to examine what were the spatial frequencies used in general, regardless of affective or contextual values. We then contrasted the classification vectors for all emotional objects and all neutral objects, and the ones for all contextual objects and all non-contextual objects, to assess the main effects of contextual and affective values. Next, we examined if there was an interaction between these two dimensions. To do so, we contrasted classification vectors of all four subcategories of objects by applying the following formula:

$$(A_1 B_1 - A_1 B_2) - (A_2 B_1 - A_2 B_2),$$

where A represents emotional value, B represents contextual value, and the number represents the level of the variable. We finally investigated the simple effects by comparing the conditions pairwise. The statistical significance of the resulting classification vectors was assessed by applying the Cluster test (Chauvin et al., 2005). Given an arbitrary z-score threshold, this test gives a cluster size above which the specified p-value is satisfied. We used this test rather than the Pixel test (Chauvin et al., 2005) because it is in general more sensitive, allowing us to detect weaker but more diffuse signals. Here, we used a threshold of ± 3 ($p < 0.05$, two-tailed). We report the size k of the significant cluster and its maximum Z-score Z_{max}. We implemented the Cluster tests as bootstraps (Efron and Tibshirani, 1993); that is, we repeated all regressions 10,000 times pairing the sampling vectors with transformed RTs randomly selected in the observed transformed RT distribution. This resulted in 10,000 random classification vectors per condition. We used these random classification vectors to transform the elements of the observed classification vectors into z-scores and estimate their p-values. We corrected p-values for multiple comparisons in the pairwise comparisons by implementing Hochberg's step-up procedure (Hochberg, 1988).

RESULTS

EFFECTS OF CONDITION AND SPATIAL FREQUENCIES ON ACCURACY

The mean accuracy was 87.49% ($SD = 7.63$). To analyse possible effects of condition on accuracy, without taking SFs into account, we first conducted a 2 (Context: non-contextual or contextual) × 2 (Emotion: neutral or emotional) repeated-measures ANOVA on mean accuracies per participant. There was an effect of contextual value [$F_{(1, 46)} = 39.83$, $p < 0.001$, $\eta_p^2 = 0.46$]: non-contextual objects ($M = 81.92\%$; $SD = 9.21$) were recognized more easily than contextual ones ($M = 77.19\%$; $SD = 10.96$). There also was an effect of emotional value [$F_{(1, 46)} = 6.31$, $p < 0.05$, $\eta_p^2 = 0.12$]: neutral objects ($M = 80.30\%$; $SD = 9.48$) were recognized slightly more easily than emotional objects ($M = 78.81\%$; $SD = 10.49$).

There was an interaction between emotional and contextual values [$F_{(1, 46)} = 53.04$, $p < 0.001$, $\eta_p^2 = 0.53$]. This interaction was decomposed into simple effects. First, there was an effect of emotion on non-contextual objects [$F_{(1, 46)} = 49.63$, $p < 0.001$, $\eta_p^2 = 0.52$]. Non-contextual neutral objects ($M = 85.58\%$; $SD = 7.94$) were recognized more easily than non-contextual emotional objects ($M = 78.26\%$; $SD = 11.49$). Second, there was an effect of emotion on contextual objects as well [$F_{(1, 46)} = 20.87$, $p < 0.001$, $\eta_p^2 = 0.31$]. Contextual emotional objects ($M = 79.36\%$; $SD = 10.31$) were recognized more easily than neutral contextual objects ($M = 75.02\%$; $SD = 12.45$).

Accuracy did not correlate significantly with the presentation of any SF.

EFFECT OF CONDITION ON RESPONSE TIMES

The mean RT for correct positive trials was 623 ms ($SD = 83$). To analyse possible effects of condition on RTs, without taking SFs into account, we conducted a 2 (Context: non-contextual or contextual) × 2 (Emotion: neutral or emotional) repeated-measures ANOVA on $-\log(x + 1)$-transformed RT means per participant (Ratcliff, 1993). Aberrant scores (over 2 s) were excluded from the analysis. There was an effect of contextual value on RTs [$F_{(1, 46)} = 161.29$, $p < 0.001$, $\eta_p^2 = 0.78$] whereby non-contextual objects (Md[1] $= 596$ ms; $SD = 60$) were recognized faster than contextual ones ($Md = 537$ ms; $SD = 67$). There was no effect of emotional value [$F_{(1, 46)} < 1$].

There also was an interaction between emotional value and contextual value [$F_{(1, 46)} = 18.46$, $p < 0.001$, $\eta_p^2 = 0.29$]. This interaction was decomposed into simple effects. First, there was an effect of emotion on non-contextual objects [$F_{(1, 46)} = 12.53$, $p < 0.001$, $\eta_p^2 = 0.21$]. Non-contextual neutral objects ($Md = 532$ ms; $SD = 57$) were identified faster than non-contextual emotional objects ($Md = 548$ ms; $SD = 68$). There also was an effect of emotion on contextual objects [$F_{(1, 46)} = 10.15$, $p < 0.01$, $\eta_p^2 = 0.18$]. Contextual emotional objects ($Md = 579$ ms; $SD = 64$) were identified faster than contextual neutral ones ($Md = 609$ ms; $SD = 80$).

---

[1]Median reaction times are given, since the ANOVA was performed on log transformed values. Given that the mean log values wouldn't be readily interpretable and that the median values don't change with a log transformation, we made this choice for purposes of clarity and transparency.

EFFECT OF SPATIAL FREQUENCIES ON RESPONSE TIME

To determine the spatial frequencies that contributed most to fast object recognition for each condition, we performed least-square multiple linear regressions between z-scored transformed RTs (see Methods: Spatial Frequency Data Analysis) and corresponding sampling vectors for correct positive trials. All object categories confounded, SFs between 13.71 and 24.31 cycles per object width (cpo) correlated negatively with RTs (peak at 19.45 cpo, $Z_{max} = 3.94$, $k = 23$, $p < 0.01$; **Figure 3A**). In other words, RTs were consistently reduced with the presentation of SFs within these boundaries. To examine a possible effect of emotional value, we contrasted classification vectors for all emotional objects and all neutral objects. There was no significant difference ($p > 0.05$). Similarly, there was no significant difference between non-contextual and contextual objects ($p > 0.05$).

We then examined the interaction between affective and contextual values (see Methods: Spatial frequency data analysis). We found a significant interaction for SFs between 5.52 and 6.69 cpo (peak at 6.02 cpo, $Z_{max} = 3.29$, $k = 3$, $p < 0.05$; **Figure 3B**).

We subsequently decomposed the interaction into simple effects. There was a significant effect of contextual value on neutral objects between 15.25 and 19.20 cpo; these SFs were correlated more negatively with RTs for contextual neutral objects than for non-contextual neutral objects (peak at 18.98 cpo, $Z_{max} = 3.36$, $k = 9$, $p < 0.05$, corrected for multiple comparisons; **Figure 3C**). However, the interaction was not significant for these SFs, making this effect difficult to interpret. There

also was an effect of contextual value for emotional objects: SFs between 4.86 and 6.56 cpo correlated more positively with RTs for contextual emotional objects than for non-contextual emotional objects (peak at 5.56 cpo, $Z_{max} = 3.75$, $k = 4$, $p < 0.05$, corrected for multiple comparisons; **Figure 3D**). Moreover, there was an effect of emotional value on contextual objects: SFs between 4.86 and 6.09 cpo correlated more positively with RTs for contextual emotional objects than for contextual neutral objects (peak at 5.56 cpo, $Z_{max} = 3.21$, $k = 3$, $p < 0.05$, corrected for multiple comparisons; **Figure 3E**). Finally, we observed no significant difference between non-contextual neutral and non-contextual emotional objects ($p > 0.05$). The interaction thus seems to be caused by the significant effect of contextual value on emotional but not on neutral objects, combined with the significant effect of emotional value on contextual but not on non-contextual objects.

DISCUSSION

GENERAL SPATIAL FREQUENCY USE

A few studies have examined the effect of specific SF band filtering during name-picture verification tasks, similar to ours. Collin and McMullen (2005) reported that low-pass filtering objects had little impact on basic-level verification (e.g., verify the "cat" category instead of the "animal" category at the superordinate level or the "Siamese" category at the subordinate level), suggesting that basic-level categorization does not especially rely on LSFs. Furthermore, Harel and Bentin (2009) reported that basic-level categorization was equally impaired by removal of either HSFs or LSFs, thus suggesting that neither of these bands is especially
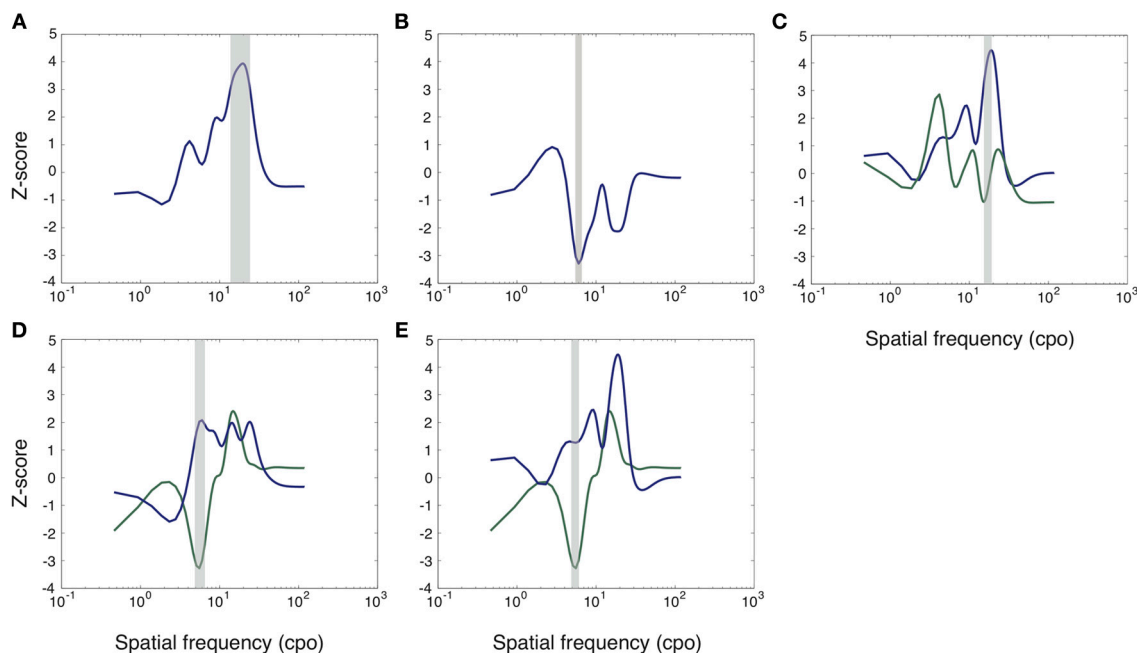

FIGURE 3 | Group classification vectors depicting the correlations between SFs and RTs for different conditions. Higher z-scores indicate a negative correlation (SFs leading to shorter RTs) while lower z-scores indicate a positive correlation (SFs leading to longer RTs). Highlighted gray areas are significant ($p < 0.05$). See text for details. **(A)** All objects together. **(B)** The vector depicting potential interactions between both variables, obtained by contrasting the contrasts of contextual value for both levels of emotional value. **(C)** Non-contextual neutral (green) objects and contextual neutral (blue) objects. **(D)** Contextual emotional (green) and non-contextual emotional (blue) objects. **(E)** Contextual emotional (green) and contextual neutral (blue) objects.

useful for recognition at the basic level. However, Harel and Bentin's cutoff for HSFs was especially high (65 cpo, or 6.5 cpd), thus preserving only very fine information typically not useful for object recognition. A large band of intermediate spatial frequencies was not explored in these studies.

An important aspect of our study is that instead of applying filters with fixed arbitrary cut-offs, we randomly sampled the entire SF spectrum. This allowed us to overcome the need of selecting arbitrary SF bands to evaluate. Indeed, there is no consensus in the literature about what consists of LSFs or HSFs: this seems to be more understood as a relative measure for SF bands inside a given study. Cut-offs for LSFs in the literature vary from 5 cpo (Boutet et al., 2003) to 15 cpo (Alorda et al., 2007). Similarly, cut-offs for HSFs vary from 20 cpo (Boutet et al., 2003) to 65 cpo (Harel and Bentin, 2009). When cut-offs are translated into cycles per degree (cpd), acknowledging that the diagnostic SFs may vary according to viewing distance, the discrepancy is even larger: cut-offs for LSFs vary from less than 0.4 cpd (Boutet et al., 2003) to more than 2.4 cpd (Alorda et al., 2007) and cut-offs for HSFs vary from 1.4 cpd (Boutet et al., 2003) to 6.5 cpd (Harel and Bentin, 2009). Quite interestingly, we note that some SFs (between 1.4 and 2.4 cpd) may be included either in LSFs or HSFs.

Our random sampling of the entire SF spectrum allowed us to evaluate the use of SFs considered as neither low nor high by most previous studies. Using this unbiased experimental approach, we found that intermediate SFs between about 14 and 24 cpo (2.3–4 cpd) are associated with fast RTs for basic-level verification. This suggests that objects are processed particularly rapidly through these SFs. Although this interpretation is the most straightforward, it is also possible that object processing was at least partly completed before the presentation of the words and, therefore, that the RTs reflect remnants of object processing rather than object processing *per se.*

Another unique aspect of our study is the fact that we equalized SF content of the object images prior to their sampling. This allows us to interpret results more confidently in terms of content of internal representations. Indeed, if SF content is not normalized among stimuli, results most likely reflect an interaction of the stored representation with the information available in the stimulus. Unfortunately, few studies have applied this procedure. As a notable exception, Willenbockel et al. (2010b) did equalize SF spectrum and randomly sample SFs in a face recognition task. Results revealed that SFs peaking at approximately 9 and 13 cycles/face (equivalent to 1.4 and 2 cpd, i.e., SFs that may be categorized as LSFs, HSFs, or most often neither of these) were most correlated with fast and accurate face identification. Although these SFs specific for images of faces are likely to differ from the SF content of object representations, they are an additional indicator that, as in the present study, intermediate SFs rather than LSFs occupy the greatest place in our representation of the world. It is plausible that stored representations consist of mostly these SFs because they are part of the intermediate band of SFs to which we are naturally most sensitive (e.g., Watson and Ahumada, 2005).

INTERACTION BETWEEN AFFECTIVE AND CONTEXTUAL VALUES

No main effect of contextual or affective value was observed in the SFs correlating with the objects' fast identification. However,

we found a significant interaction between affective and contextual values for SFs centered on 6 cpo (or 1 cpd). This indicates that these LSFs, those usually associated with the magnocellular pathway (Derrington and Lennie, 1984), are sensitive in a nonlinear manner to a combination of the visual object's intrinsic properties.

When testing the simple effects, we observed that affective value elicited a significant difference in the use of these SFs in contextual objects: they led to longer RTs for contextual emotional objects than for contextual neutral ones. This is not in accordance with the general effect of affective value usually reported in the literature (i.e., LSFs leading to faster RTs, e.g., Mermillod et al., 2010); however, our result is due to an interaction between affective and contextual values and is therefore difficult to compare to those of other studies. Moreover, our stimuli were equalized in their SF content and always comprised several randomly sampled SF bandwidths at the same time, whereas in studies using filters with fixed cut-offs, only some specific band of LSFs or HSFs is shown at a time.

SFs near 6 cpo (or 1 cpd) also led to longer RTs for contextual emotional objects than for non-contextual emotional objects. The effect of contextual value on SF content of object representations had not been tested before but it had been often proposed that rapidly extracted LSFs are sufficient to activate representations associated with an object (Bar, 2004; Fenske et al., 2006). Our data suggest that these presumed/hypothetical associative representations do not speed up the object's recognition. Why we observed this modulation only for emotional objects is not clear, but several interactions between affective and contextual processing have already been reported and could possibly explain the discrepancy (Storbeck and Clore, 2005; Brunyé et al., 2013; Shenhav et al., 2013). For example, affective value might influence the extent to which we associate a particular object to other objects (Bar, 2009; Shenhav et al., 2013).

CONCLUSION

The main findings of the present study are (i) that the SF content of object representations in general are in an intermediate band between 14 and 24 cpo (2.3–4 cpd), and (ii) that intrinsic high-level categorical properties of an object influence the SF content of its internally stored representation, more precisely that affective and contextual values interact in their modulation of the SF content of object representations.

According to predictive accounts of brain function (e.g., Rao and Ballard, 1999; Bar, 2003; Friston, 2003, 2010; Friston et al., 2006), our mind constantly generates predictions about our environment, and our understanding of a sensory input is based both on the available sensory information and on prior beliefs stored as internal representations (see Knill and Pouget, 2004). In this study, we investigated precisely the SF content of these stored representations, and its potential flexible modulation by affective and contextual properties of the stimulus. Our results reveal that stored representations of visual objects are composed of intermediate SFs that are often left over in studies using filters with fixed arbitrary cut-offs. Furthermore, we observed a modulation of this SF content by affective and contextual intrinsic values of the visual object, suggesting its flexibility and thus the multiplicity of visual recognition systems.

Our study cannot however address directly the issue of temporal dynamics of visual object recognition. While we observed that some SFs are more useful to identify some objects, we cannot conclude that these are extracted first. Further studies should address these issues and their links to potential initiation of top-down mechanisms.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fpsyg.2014. 00512/abstract

REFERENCES

Alorda, C., Serrano-Pedraza, I., Campos-Bueno, J. J., Sierra-Vázquez, V., and Montoya, P. (2007). Low spatial frequency filtering modulates early brain processing of affective complex pictures. *Neuropsychologia* 45, 3223–3233. doi: 10.1016/j.neuropsychologia.2007.06.017

Aminoff, E., Gronau, N., and Bar, M. (2007). The parahippocampal cortex mediates spatial and nonspatial associations. *Cereb. Cortex* 17, 1493–1503. doi: 10.1093/cercor/bhl078

Barceló, F., Suwazono, S., and Knight, R. T. (2000). Prefrontal modulation of visual processing in humans. *Nat. Neurosci.* 3, 399–403. doi: 10.1038/73975

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *J. Cogn. Neurosci.* 15, 600–609. doi: 10.1162/0898929033216 62976

Bar, M. (2004). Visual objects in context. *Nat. Rev. Neurosci.* 5, 617–629. doi: 10.1038/nrn1476

Bar, M. (2009). Predictions: a universal principle in the operation of the human brain. *Philos. Trans. R. Soc. B Biol. Sci.* 364, 1181–1182. doi: 10.1098/rstb. 2008.0321

Bar, M., and Aminoff, E. (2003). Cortical analysis of visual context. *Neuron* 38, 347–358. doi: 10.1016/S0896-6273(03)00167-3

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., et al. (2006). Top-down facilitation of visual recognition. *Proc. Natl. Acad. Sci. U.S.A.* 103, 449–454. doi: 10.1073/pnas.0507062103

Bar, M., and Ullman, S. (1996). Spatial context in recognition. *Perception* 25, 343–352. doi: 10.1068/p250343

Barrett, L. F., and Bar, M. (2009). See it with feeling: affective predictions during object perception. *Philos. Trans. R. Soc. B Biol. Sci.* 364, 1325–1334. doi: 10.1093/brain/106.2.473

Biederman, I. (1972). Perceiving real-world scenes. *Science* 177, 77–80. doi: 10.1126/science.177.4043.77

Biederman, I. (1981). Do background depth gradients facilitate object identification? *Perception* 10, 573–578. doi: 10.1068/p100573

Biederman, I., Mezzanotte, R. J., and Rabinowitz, J. C. (1982). Scene perception: detecting and judging objects undergoing relational violations. *Cogn. Psychol.* 14, 143–177. doi: 10.1016/0010-0285(82)90007-X

Bocanegra, B. R., and Zeelenberg, R. (2009). Emotion improves and impairs early vision. *Psychol. Sci.* 20, 707–713. doi: 10.1111/j.1467-9280.2009.02354.x

Boutet, I., Collin, C., and Faubert, J. (2003). Configural face encoding and spatial frequency information. *Percept. Psychophys.* 65, 1078–1093. doi: 10.3758/BF03194835

Bradley, M. M. (2009). Natural selective attention: orienting and emotion. *Psychophysiology* 46, 1–11. doi: 10.1111/j.1469-8986.2008.00702.x

Brainard, D. H. (1997). The psychophysics toolbox. *Spat. Vis.* 10, 433–436. doi: 10.1163/156856897X00357

Bredfeldt, C. E., and Ringach, D. L. (2002). Dynamics of spatial frequency tuning in macaque V1. *J. Neurosci.* 22, 1976–1984.

Brunyé, T. T., Gagnon, S. A., Paczynski, M., Shenhav, A., Mahoney, C. R., and Taylor, H. A. (2013). Happiness by association: breadth of free association influences affective states. *Cognition* 127, 93–98. doi: 10.1016/j.cognition.2012. 11.015

Bullier, J. (2001). Integrated model of visual processing. *Brain Res. Rev.* 36, 96–107. doi: 10.1016/S0165-0173(01)00085-6

Calderone, D. J., Hoptman, M. J., Martínez, A., Nair-Collins, S., Mauro, C. J., Bar, M., et al. (2013). Contributions of low and high spatial frequency processing to impaired object recognition circuitry in schizophrenia. *Cereb. Cortex* 23, 1849–1858. doi: 10.1093/cercor/bhs169

Chauvin, A., Worsley, K. J., Schyns, P. G., Arguin, M., and Gosselin, F. (2005). Accurate statistical tests for smooth classification images. *J. Vis.* 5, 659–667. doi: 10.1167/5.9.1

Collin, C. A., and McMullen, P. A. (2005). Subordinate-level categorisation relies on high spatial frequencies to a greater degree than basic-level categorisation. *Percept. Psychophys.* 67, 354–364. doi: 10.3758/BF03206498

Derrington, A. M., and Lennie, P. (1984). Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque. *J. Physiol.* 357, 219–240.

De Valois, R. L., and De Valois, K. K. (1990). *Spatial Vision*. New York, NY: Oxford University Press.

Efron, B., and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. London: Chapman and Hall/CRC.

Fenske, M. J., Aminoff, E., Gronau, N., and Bar, M. (2006). Top-down facilitation of visual object recognition: object-based and context-based contributions. *Prog. Brain Res.* 155, 3–21. doi: 10.1016/S0079-6123(06)55001-0

Friston, K. (2003). Learning and inference in the brain. *Neural Netw.* 16, 1325–1352. doi: 10.1016/j.neunet.2003.06.005

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 607–608. doi: 10.1038/nrn2787-c2

Friston, K., Kilner, J., and Harrison, L. (2006). A free energy principle for the brain. *J. Physiol.* 100, 70–87. doi: 10.1016/j.jphysparis.2006.10.001

Gilbert, C. D., and Sigman, M. (2007). Brain states: top-down influences in sensory processing. *Neuron* 54, 677–696. doi: 10.1016/j.neuron.2007.05.019

Grossberg, S. (1980). Biological competition: decision rules, pattern formation, and oscillations. *Proc. Natl. Acad. Sci. U.S.A.* 77, 2338–2342. doi: 10.1073/pnas.77.4.2338

Harel, A., and Bentin, S. (2009). Stimulus type, level of categorization, and spatial-frequencies utilization: implications for perceptual categorization hierarchies. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 1264–1273. doi: 10.1037/a0013621

Hegdé, J. (2008). Time course of visual perception: coarse-to-fine processing and beyond. *Prog. Neurobiol.* 84, 405–439. doi: 10.1016/j.pneurobio.2007.09.001

Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* 75, 800–802. doi: 10.1093/biomet/75.4.800

Knill, D. C., and Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci.* 27, 712–719. doi: 10.1016/j.tins.2004.10.007

Kveraga, K., Boshyan, J., and Bar, M. (2007). Magnocellular projections as the trigger of top-down facilitation in recognition. *J. Neurosci.* 27, 13232–13240. doi: 10.1523/JNEUROSCI.3481-07.2007

Lebrecht, S., Bar, M., Barrett, L. F., and Tarr, M. J. (2012). Micro-valences: perceiving affective valence in everyday objects. *Front. Psychol.* 3:107. doi: 10.3389/fpsyg.2012.00107

Maunsell, J., and Newsome, W. T. (1987). Visual processing in monkey extrastriate cortex. *Annu. Rev. Neurosci.* 10, 363–401. doi: 10.1146/annurev.ne.10.030187. 002051

Mermillod, M., Droit-Volet, S., Devaux, D., Schaefer, A., and Vermeulen, N. (2010). Are coarse scales sufficient for fast detection of visual threat? *Psychol. Sci.* 21, 1429–1437. doi: 10.1177/0956797610381503

Mermillod, M., Guyader, N., and Chauvin, A. (2005). The coarse-to-fine hypothesis revisited: evidence from neuro-computational modeling. *Brain Cogn.* 57, 151–157. doi: 10.1016/j.bandc.2004.08.035

Musel, B., Chauvin, A., Guyader, N., Chokron, S., and Peyrin, C. (2012). Is coarse-to-fine strategy sensitive to normal aging? *PLoS ONE* 7:e38493. doi: 10.1371/journal.pone.0038493

Logothetis, N. K., and Sheinberg, D. L. (1996). Visual object recognition. *Annu. Rev. Neurosci.* 19, 577–621.

Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Mem. Cognit.* 3, 519–526. doi: 10.3758/BF03197524

Pascual-Leone, A., and Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science* 292, 510–512. doi: 10.1126/science.1057099

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* 10, 437–442. doi: 10.1163/15685689 7X00366

Rao, R. P., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87. doi: 10.1038/4580

Ratcliff, R. (1993). Methods for dealing with reaction outliers. *Psychol. Bull.* 114, 510–532. doi: 10.1037/0033-2909.114.3.510

Schyns, P. G., and Oliva, A. (1994). From blobs to boundary edges: evidence for time- and spatial-scale-dpendent scene recognition. *Psychol. Sci.* 5, 195–200. doi: 10.1111/j.1467-9280.1994.tb00500.x

Shenhav, A., Barrett, L., and Bar, M. (2013). Affective value and associative processing share a cortical substrate. *Cogn. Affect. Behav. Neurosci.* 13, 46–59. doi: 10.3758/s13415-012-0128-4

Storbeck, J., and Clore, G. L. (2005). With sadness comes accuracy; with happiness, false memory: mood and the false memory effect. *Psychol. Sci.* 16, 785–791. doi: 10.1111/j.1467-9280.2005.01615.x

Tomita, H., Ohbayashi, M., Nakahara, K., Hasegawa, I., and Miyashita, Y. (1999). Top-down from prefrontal cortex in executive control of memory retrieval. *Nature* 401, 699–703. doi: 10.1038/44372

Ullman, S. (1995). Sequence seeking and counter streams: a computational model for bidirectional information flow in the visual cortex. *Cereb. Cortex* 5, 1–11. doi: 10.1093/cercor/5.1.1

Vuilleumier, P., Armony, J. L., Driver, J., and Dolan, R. J. (2003). Distinct spatial frequency sensitivities for processing faces and emotional expressions. *Nat. Neurosci.* 6, 624–631. doi: 10.1038/nn1057

Watson, A. B., and Ahumada, A. J. (2005). A standard model for foveal detection of spatial contrast. *J. Vis.* 5, 6–6. doi: 10.1167/5.9.6

Watt, R. J. (1987). Scanning from coarse to fine spatial scales in the human visual system after the onset of a stimulus. *J. Opt. Soc. Am. A* 4, 2006–2021. doi: 10.1364/JOSAA.4.002006

Willenbockel, V., Fiset, D., Chauvin, A., Blais, C., Arguin, M., Tanaka, J. W., et al. (2010b). Does face inversion change spatial frequency tuning? *J. Exp. Psychol. Hum. Percept. Perform.* 36, 122–135. doi: 10.1037/a0016465

Willenbockel, V., Lepore, F., Nguyen, D. K., Bouthillier, A., and Gosselin, F. (2012). Spatial frequency tuning during the conscious and non-conscious perception of emotional facial expressions—an intracranial ERP study. *Front. Psychol.* 3:237. doi: 10.3389/fpsyg.2012.00237

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., and Tanaka, J. W. (2010a). Controlling low-level image properties: The SHINE toolbox. *Behav. Res. Methods* 42, 671–684. doi: 10.3758/BRM.42.3.671