Author: Charles S. Burlingham

Advisors: Marlene Behrmann & Elissa Aminoff

April 15, 2015

Senior Honors Thesis:

Role of spatial frequency in rapid scene processing

Acknowledgments

Thanks to the Small Undergraduate Research Grant, the Undergraduate Research Office, and the Dietrich Senior Honors Thesis Program at Carnegie Mellon University for supporting this research. Much thanks to my mentor and advisor Marlene Behrmann for guiding me through every step of this research and supporting good practices throughout my undergraduate career. Also thanks to Elissa Aminoff, the other advisor on this project, whose expertise in scene processing and general guidance in all areas has been invaluable. Thanks to Kevin Tan and Madeleine Varner for helping me fix and troubleshoot the experiment's code.

Abstract

Rapid processing of scenes has been shown to be facilitated by the processing of low spatial frequency (LSF) components before that of high spatial frequency (HSF) components. It has been been proposed that magnocellular pathways carrying low spatial frequencies provide an initial 'gist' of a scene, which is filled in using higher spatial frequencies along the the ventral stream. This study seeks to determine how performance on a six-way scene categorization task varies as a function of spatial frequency and image presentation duration. Across six categories of scenes, a low-high continuum of spatial frequencies, and a short 50ms presentation duration and longer 100ms one, we measured subjects reaction times and accuracies in categorizing scenes. Accuracy was equivalent for HSF-filtered scenes and LSF-filtered and this was true for the shorter and longer duration conditions. However, accuracy was significantly better for HSF-filtered images of forests at 100ms versus at 50ms presentation. Taken together, these results suggest that in addition to LSF information, HSF components of scenes may also facilitate rapid scene processing for some types of scenes.

Introduction

Humans are able to rapidly perceive and recognize visual scenes in roughly the same amount of time as it takes to recognize an object such as a car, face, or dog (Olivia, 2013). Potter and Levy's early behavioral experiments revealed that humans are able to understand, remember, and describe complex real-world images after seeing them only for 100ms (Potter & Levy 1969). Thorpe et al.'s ERP study revealed that visual processing of objects within complex real-world images can be performed in under 150ms (Thorpe et al. 1996). Olivia and Schyns' foundational visual psychophysics study on rapid scene processing showed that humans can recognize complex visual scenes in 125ms (Oliva & Schyns 1994).

The speed of processing shown by these studies is surprising, considering that visual scenes are complex and often contain many objects with varying degrees of contextual association. Hence, mental representations of scenes are likely significantly different from those of individual objects. Recent studies employing transcranial magnetic stimulation (TMS) have shown that, at least within the ventral visual stream, the neural mechanisms mediating object and scene processing may be functionally dissociable (Ganaden et al. 2013, Mullin & Steeves 2011).

The geometries of scenes, and in particular their global spatial information, are thought to play an important role in rapid scene processing (Bar et al. 2006, Oliva & Torralba 2006). Olivia & Torralba (2001) demonstrated with a computational model that a low dimensional representation of a scene image, termed the 'spatial envelope,' can sufficiently convey enough information to categorize scenes semantically, like humans

do. The spatial envelope, described by visual parameters such as naturalness, openness, and closeness, is a holistic representation of scenes that is not defined by the information conveyed by individual objects within it. Rather, these visual parameters are correlated well with the second-order statistics (energy spectra) of scene images as well as the spatial arrangement of structures in the scene (spectrogram) (Oliva & Torralba 2001).

Different ranges of spatial frequencies have been shown to mediate specialized mechanisms in visual recognition of objects, scenes, faces, and words (Peyrin et al. 2006, Bar et al. 2006, Vuilleumier et al. 2003, Woodhead et al. 2011). In the domain of object recognition, Bar et al. and others have posited that a top-down mechanism — namely a rapid projection from early visual cortex to the OFC and then to the ventral visual stream — facilitates the extraction of an initial 'gist' from the scene conveyed by low spatial frequencies (Bar et al. 2006, Kveraga et al. 2007).

The low spatial frequency components of scenes have been shown to be sufficient for rapid scene processing. Olivia & Schyns foundational study on rapid scene processing employed an image matching task with low, high and LF-HF hybrid scene images. They found that the processing of LSF occurs in advance of that of HSF, and is sufficient for recognition (Oliva & Schyns 1994). It has been posited that the low spatial frequencies of scenes are transmitted rapidly by a magnocellular pathway, mediating a coarse estimation or gist (Kveraga et al. 2007). This LSF-based gist is proposed to be subsequently used as a template for further processing via analysis of high spatial frequencies in the ventral visual pathway (Bar & Aminoff, 2003; Bullier, 2001; Hegdé, 2008; Kauffmann et al., 2014; Schyns & Oliva, 1994). This LSF-sensitive top-down

mechanism has been proposed to facilitate the rapidity of visual processing in general (Bar et al 2006, Bar 2007).

The current visual psychophysics experiment investigates how performance on a rapid six-way scene categorization task varies as a function of spatial frequency and presentation duration. We hypothesized that LSF components should be processed prior to HSF components of scenes, as Bar et al. (2006) found. Hence, we predicted that when the scenes were presented for the very brief 50ms duration, there would be facilitated performance for the LSF scenes, but not the HSF ones, and that there would an inverse linear relationship between SF and performance. We also predicted that at the longer 100ms duration, this facilitation would disappear, such that performance would be equivalent across the SF conditions.

Methods

Participants

Twenty-eight right-handed participants (8 males, 20 females, 20 ± 2 years) with normal or corrected-to-normal vision were included in this experiment. All participants were all undergraduate students at Carnegie Mellon University and native English speakers. Participants gave informed written consent before participating in the study, which was approved by the IRB at Carnegie Mellon University.

Stimuli

Scene stimuli were taken from multiple sources including from the image database from Park, Konkle, & Oliva (2014), from those used in Kravitz, Peng, & Baker (2011), and from freely licensed images found using Google image search (those labeled “for reuse with modification”). The six scene categories — bedrooms, churches, mountains, skylines, streams, & woods — were chosen in order to span the space of possible spatial frequencies inherent to the scene types and to represent variation in the ‘spatial envelope’ as widely as possible (Oliva & Torralba 2001). Hence, the scene categories chosen were half manmade and half natural, and within these divisions, some close and some far, and some closed and open, as Kravitz, Peng, & Baker chose to do. Exemplars from each of the six categories were included in the experiment, for a total of 120 individual scene exemplars. Bedrooms, churches, and skylines constituted the categories within the ‘manmade’ parameter, whereas mountains, streams, and forests constituted the ‘natural’ parameter. Bedrooms, streams, and forests made up the ‘near’

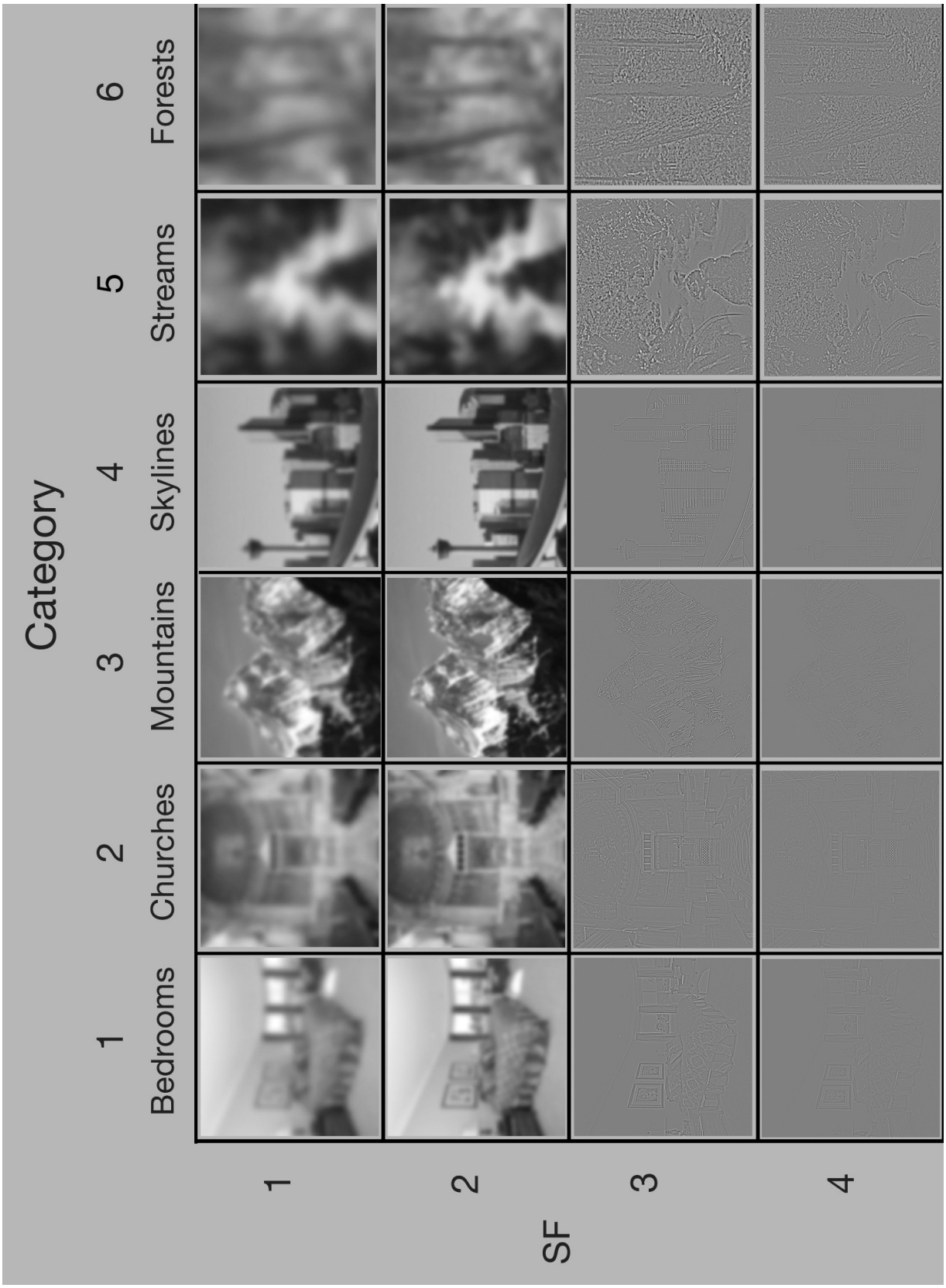

Figure 1a: One exemplar from each category of scene is shown in all four spatial frequency conditions, from low to high (1 to 4).

parameter and churches, skylines, and mountains made up the ‘far’ parameter. Bedrooms, churches, and forests made up the ‘closed’ parameter and mountains, skylines, and streams made up the ‘open’ parameter (see figure 1a).

Using Photoshop CC, all exemplar images were converted to grayscale by changing the image mode from ‘RGB’ to ‘Grayscale’. Then, all images were resized or cropped to 500px by 500px. Each of the 120 exemplar images was filtered to include primarily low spatial frequencies by running the images through the ‘Gaussian Blur’ filter using a pixel radius of 8.5. This set of 120 filtered images constituted the group ‘SF 1’, with the highest proportion of low spatial frequencies, and the lowest proportion of high spatial frequencies. To produce the group ‘SF 2’ — images with more high spatial frequencies than images in ‘SF 1,’ but still predominately low spatial frequencies — the original unfiltered exemplar images were run through the ‘Gaussian Blur’ filter using a pixel radius of 6.1. Next, each of the original unfiltered images was filtered to include primarily high spatial frequencies by running the images through the ‘High Pass’ filter with a pixel radius of 1.4. This set of images constituted the group ‘SF 3,’ with the second highest amount of high spatial frequencies, and few low spatial frequencies. To produce the group ‘SF 4’ — images with highest amount of high spatial frequencies among the four SF conditions, and few low spatial frequencies — the original unfiltered exemplar images were run through the ‘High Pass’ filter with a pixel radius of .8. Overall, 480 image files were produced, comprising the original 120 exemplar images within each of the 4 SF filtered groups. Groups SF 1, SF 2, SF 3, & SF 4 represent a continuum from mostly low spatial frequencies to mostly high spatial frequencies although at every level of spatial frequency, there are both high and low spatial frequencies present, albeit with differential weighting (see figure 1b).

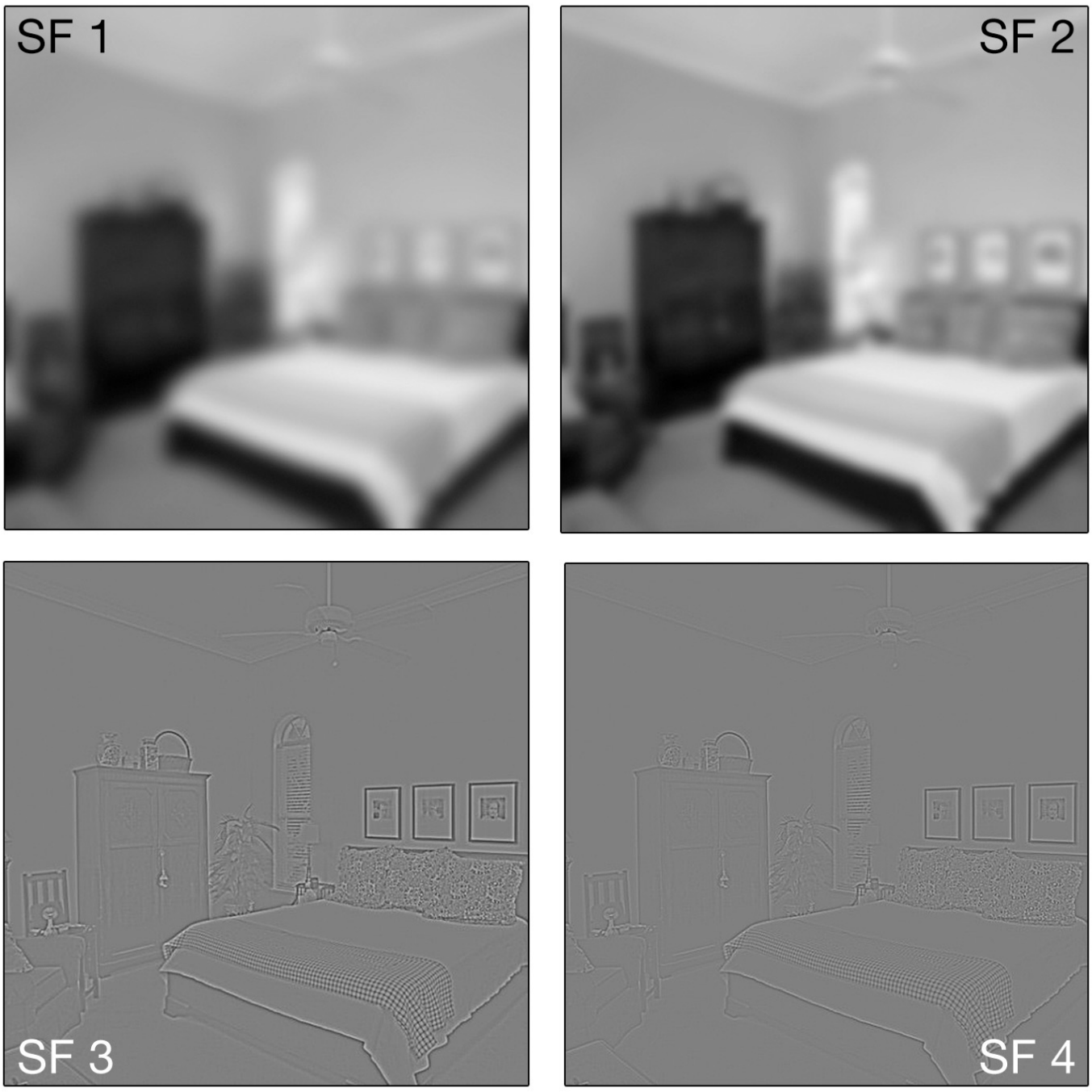

Figure 1b: An exemplar from the category ‘Bedrooms’ is depicted in all four spatial frequency conditions.

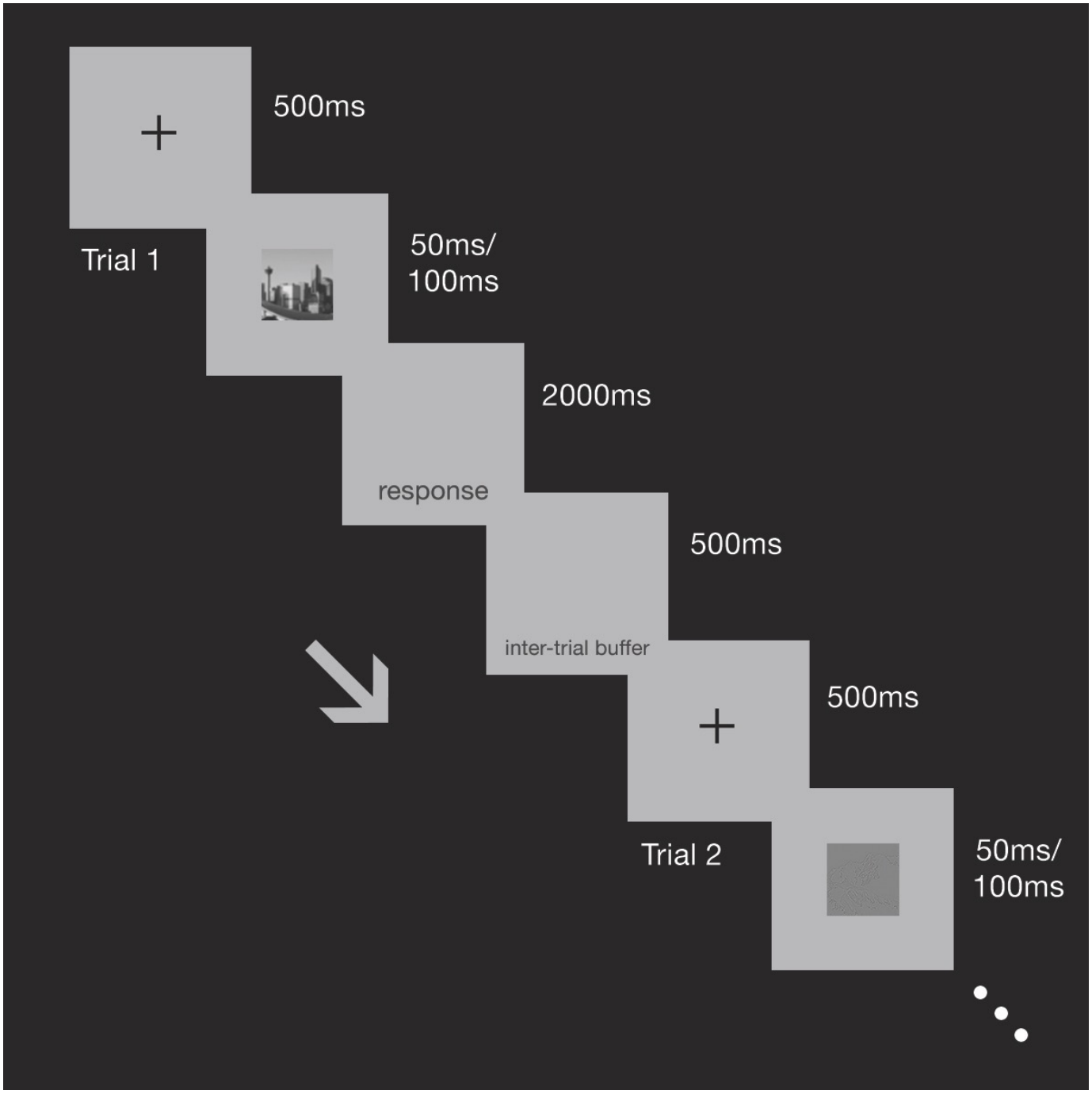

Figure 2: A single trial broken down to show the order and presentation duration of fixation, stimulus, and inter-trial buffer, as well as the time limit for subjects' response.

Experimental Design

The stimuli were presented to participants on a square LCD monitor connected via HDMI to a Macbook Pro running PsychToolBox in MATLAB. Using a setup with an external monitor was necessary in order to control precisely the duration that each stimulus was presented for (50ms or 100ms ± 6ms due to lag). The experiment was written in PsychToolBox for MATLAB, and additional Python and MATLAB scripts were used to generate individual trial orders for each subject and parse data output.

The experimental procedure was broken down into three practice blocks followed by two experimental blocks, with breaks in between each. Each trial consisted of a fixation cross presented for 500ms, then the image presented on a white background for either 50ms or 100ms, a response window capped at 2000ms, and then a post-image white screen for 500ms (see figure 2).

The first practice block consisted of 50 trials, in which filtered and unfiltered scene images were presented for 100ms, none of which were repeated in the following blocks. In the second practice block, a first set of 120 stimuli, randomized according to their category, exemplar and SF, were presented for 50ms each. In the third and final practice block, these same 120 stimuli were presented again, but their order was randomly permutated, and they were presented for 100ms each. Each of the two experimental blocks also consisted of 120 trials. In the first experimental block, a second set of 120 stimuli were presented for 50ms each. In the second experimental block, this second set was presented again for 100ms each with their order randomly permutated. Overall, therefore, subject 1 saw 240 of the total 480 images and subject 2 saw the remaining 240 images. Subject 3 saw 240 of a new set of 480 images, re-randomized according to category, exemplar and SF, and subject 4 saw the remaining 240 images of this new set.

Procedure

Participants were asked to categorize the images presented to them as belonging to one of six possible scene categories using keys number 1-6 on the keyboard. Category 1 — ‘Bedrooms’ — corresponded to key 1, category 2 — ‘Churches’ — corresponded to key 2, category 3 — ‘Mountains’ — corresponded to key 3, category 4 — ‘Skylines’ — corresponded to key 4, category 5 — ‘Streams’ — corresponded to key 5, and category 6 — ‘Woods’ — corresponded to key 6. Participants were shown these category–response associations on-screen and asked to memorize them at the beginning of the experiment. They were also able to practice categorizing the scenes during the practice block. During the four experimental blocks, if the participant took longer than 2 seconds to respond or pressed a key outside of the range of possible responses, they were prompted with feedback on-screen about their error. No other feedback was given post-trial regarding the correctness of their response. Response time and accuracy were recorded for each trial. The program also recorded the actual amount of time each image was presented as a measure of possible lag caused by the hardware/software. The experiment took most participants approximately 20 minutes to complete.

Results

In order to explore differences in performance across the four spatial frequency conditions, the six categories, and the two duration conditions, two repeated measures ANOVAs were performed separately for percent correct and RT means on just the second half of trials (trials 240-480).

The final analysis was performed just on the second half of trials because initial analyses revealed that performance was worse overall in the first half of trials (accuracy = 80%, RT = 833ms) than in the second (accuracy = 86.5%, RT = 768ms). Some subjects also reported difficulty learning the key-category associations, even after practice. Hence, there may have been unforeseen deleterious effects of performance on the first 120 trials, in which images were presented for 50ms, which could potentially bias the analysis of duration effects and cause the observed decrease in overall performance on the first half of trials. In the second half of trials, subjects have equivalent practice at both durations — 50ms and 100ms — which is crucial as rapid scene processing is the primary object of investigation, not training or response related effects.

Prior to the analyses of variance, each subject's performance was analyzed individually and those whose mean RT over all trials exceeded the mean + two standard deviations for all subjects were removed from further analyses. These criteria for removal were also applied to subjects' mean accuracy. Eight subjects of a total of thirty-six were removed, leaving twenty-eight whose data were used in the following analyses.

The first analysis, a 6 x 4 x 2 repeated measures ANOVA, examined mean RT for just those trials that subjects answered correctly on, on just the second half of trials in the experiment. There was a significant main effect of category ($F(1, 5)= 8.328$, $p < 0.001$) and a significant main effect of SF ($F(1,3) = 5.690$, $p = .004$). There were neither significant two-way interactions between the factors, nor a significant three-way interaction between category, SF, and duration (see figures 3a & 3b).

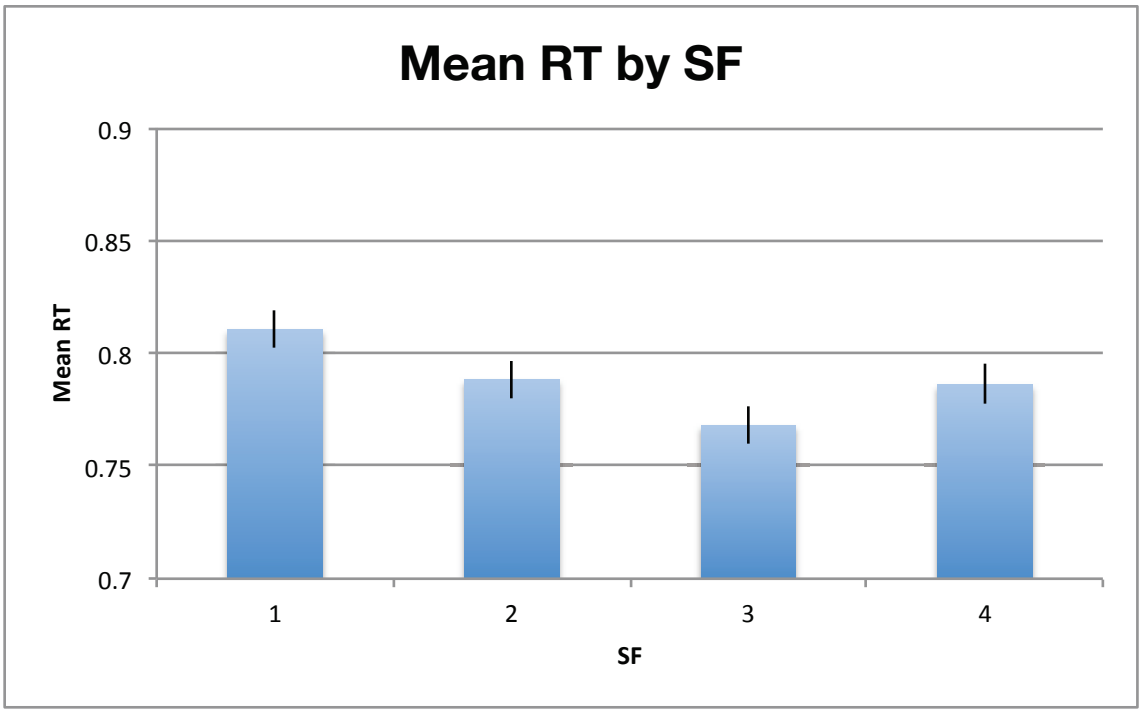

Figure 3a: The significant main effect of SF for the RT ANOVA is shown here, with fastest RT for SF 3 and slowest RT for SF 1.

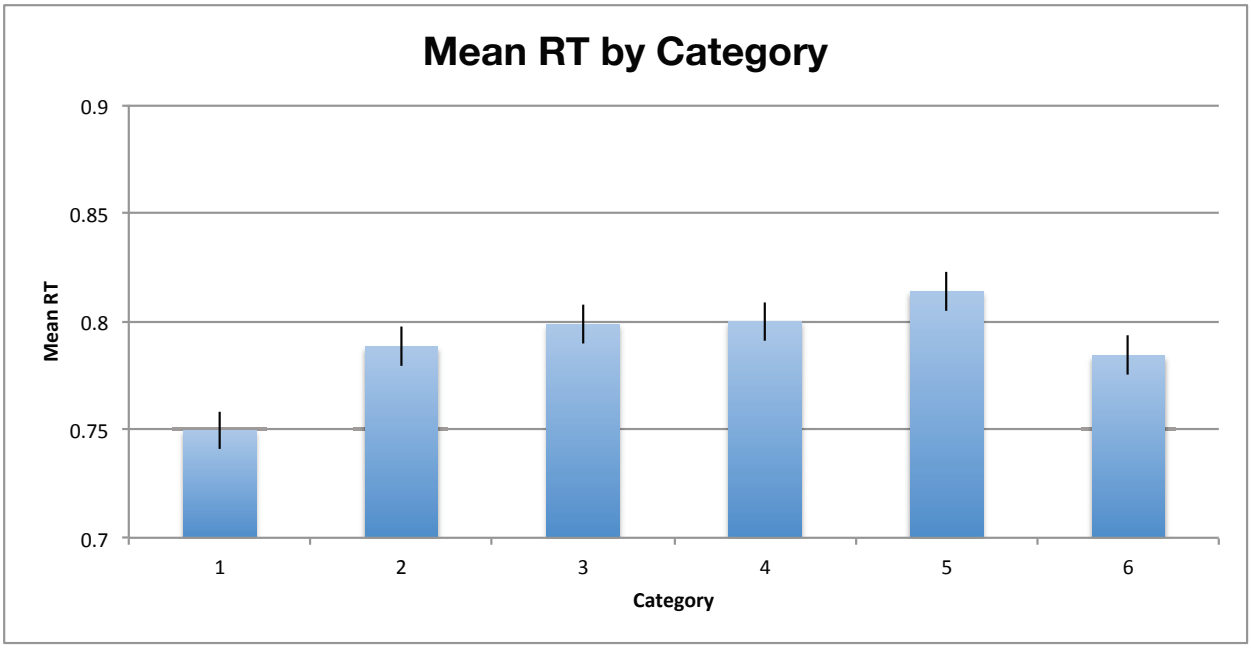

Figure 3b: The significant main effect of category for the RT ANOVA is shown here, with fastest RT for category 1, bedrooms, and slowest RT for category 5, streams.

The second ANOVA examined accuracy on just the second half of trials in the experiment. Accuracy, rather than RT, was chosen as the metric of performance based on the difficulty of the categorization and because it is a commonly used metric in perceptual tasks (Santee & Egeth 1982). There was a significant main effect of category ($F(1, 5) = 10.471$, $p < .001$) and of SF ($F(1,3) = 12.307$, $p < .001$) and of duration ($F(1,1) = 12.074$, $p = .002$) (see figures 4a, 4b, & 4c).

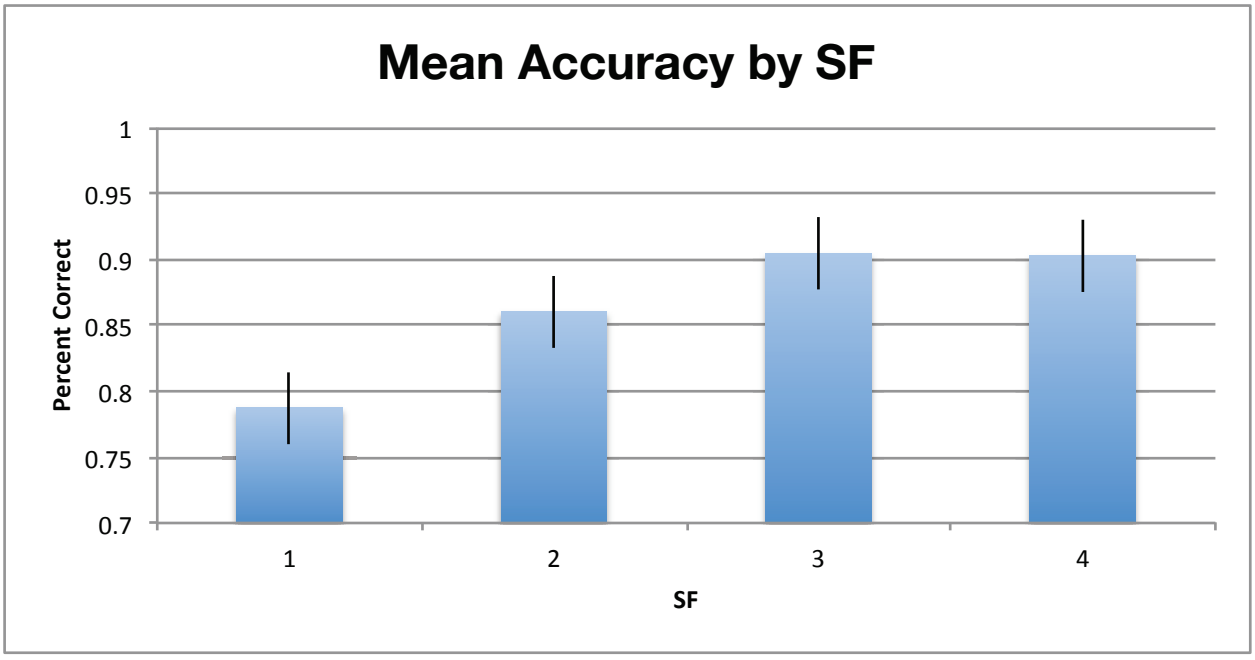

Figure 4a: Mean accuracy for the four SF conditions, where chance = 17%.

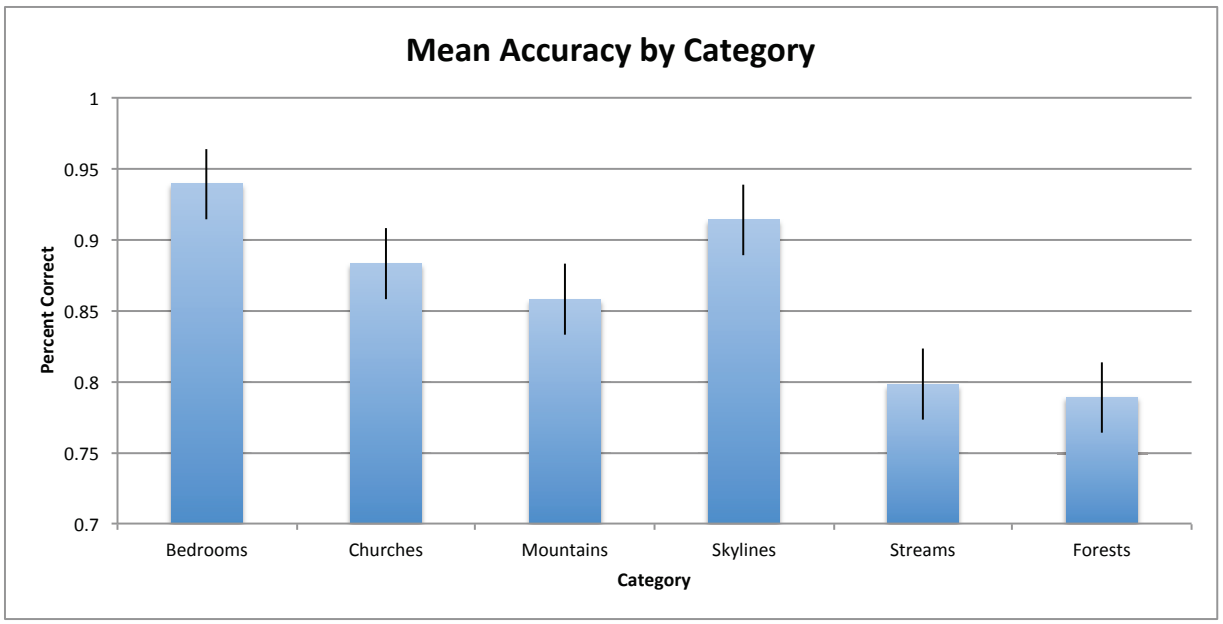

Figure 4b: Mean accuracy for the six categories, where chance = 17%.

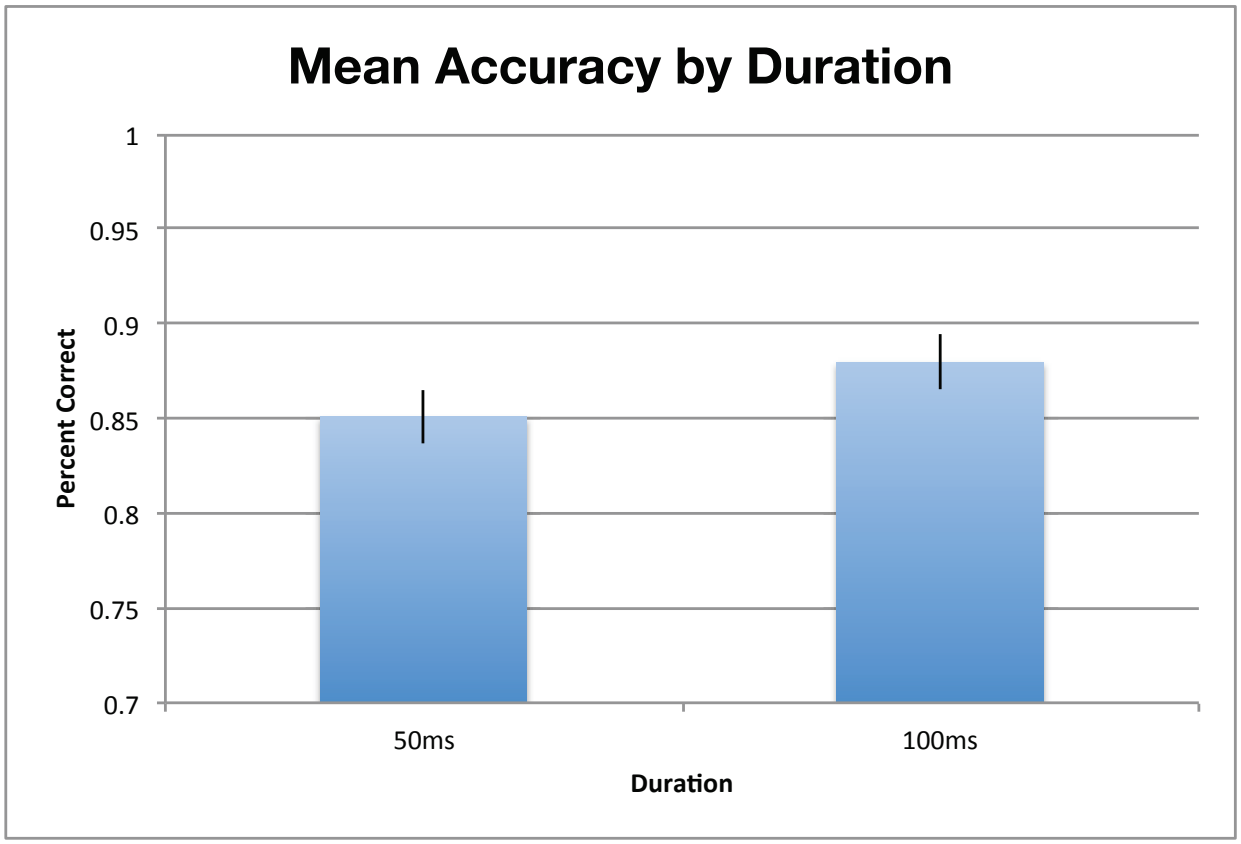

Figure 4c: Mean accuracy for the two duration conditions, where chance = 17%.

There was a significant two-way interaction between category and SF ($F(1,15) = 8.136$, $p < .001$). There was a marginally significant two-way interaction between category and duration ($F(1,5) = 2.172$, $p = .061$). The two-way interaction between SF and duration did not reach significance. There was also a significant three-way interaction between category, SF, and duration ($F(1,15) = 1.891$, $p < .023$) (see figures 4d, 4e, & 4f).

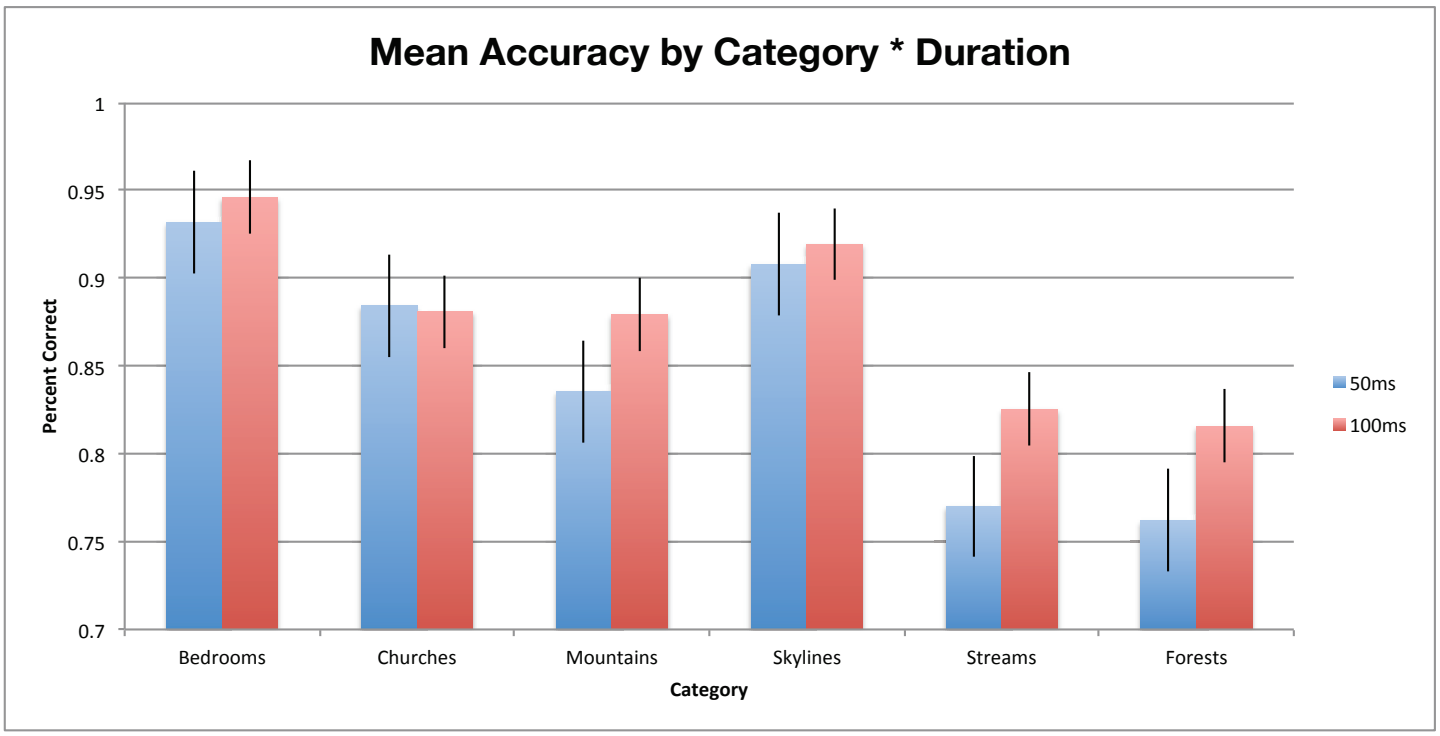

Figure 4d: Mean accuracy for six categories and two durations, where chance = 17%. Note that the two-way interaction for category and duration was only marginally significant.

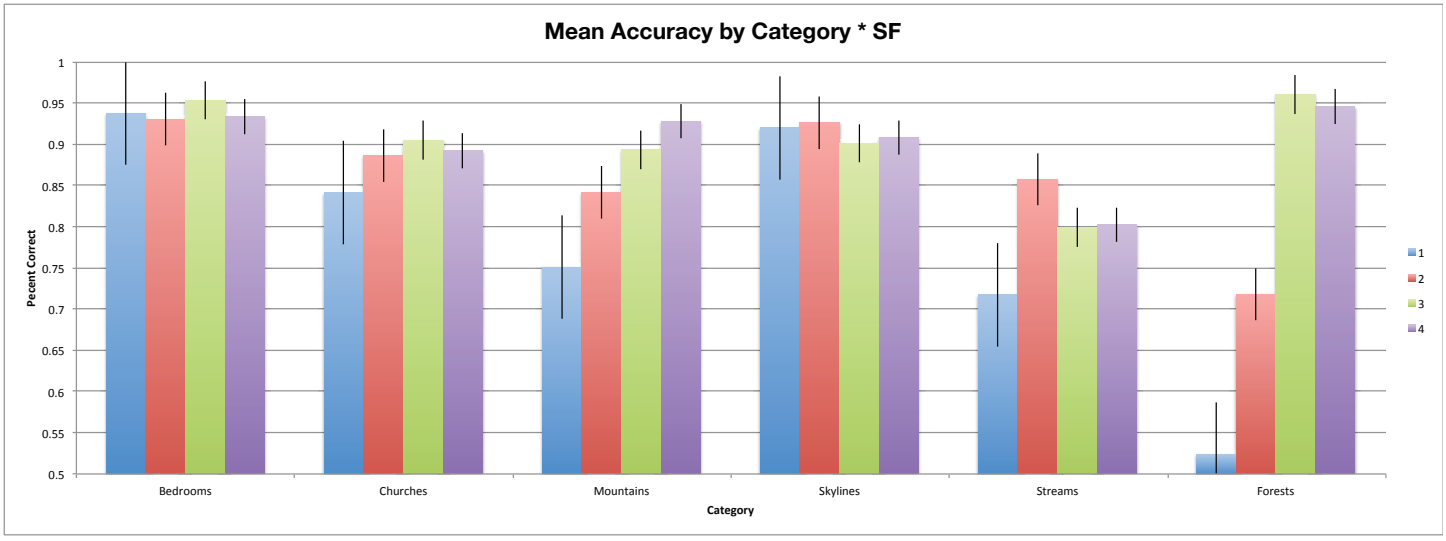

Figure 4e: Mean accuracy for six categories and four SF, where chance = 17%

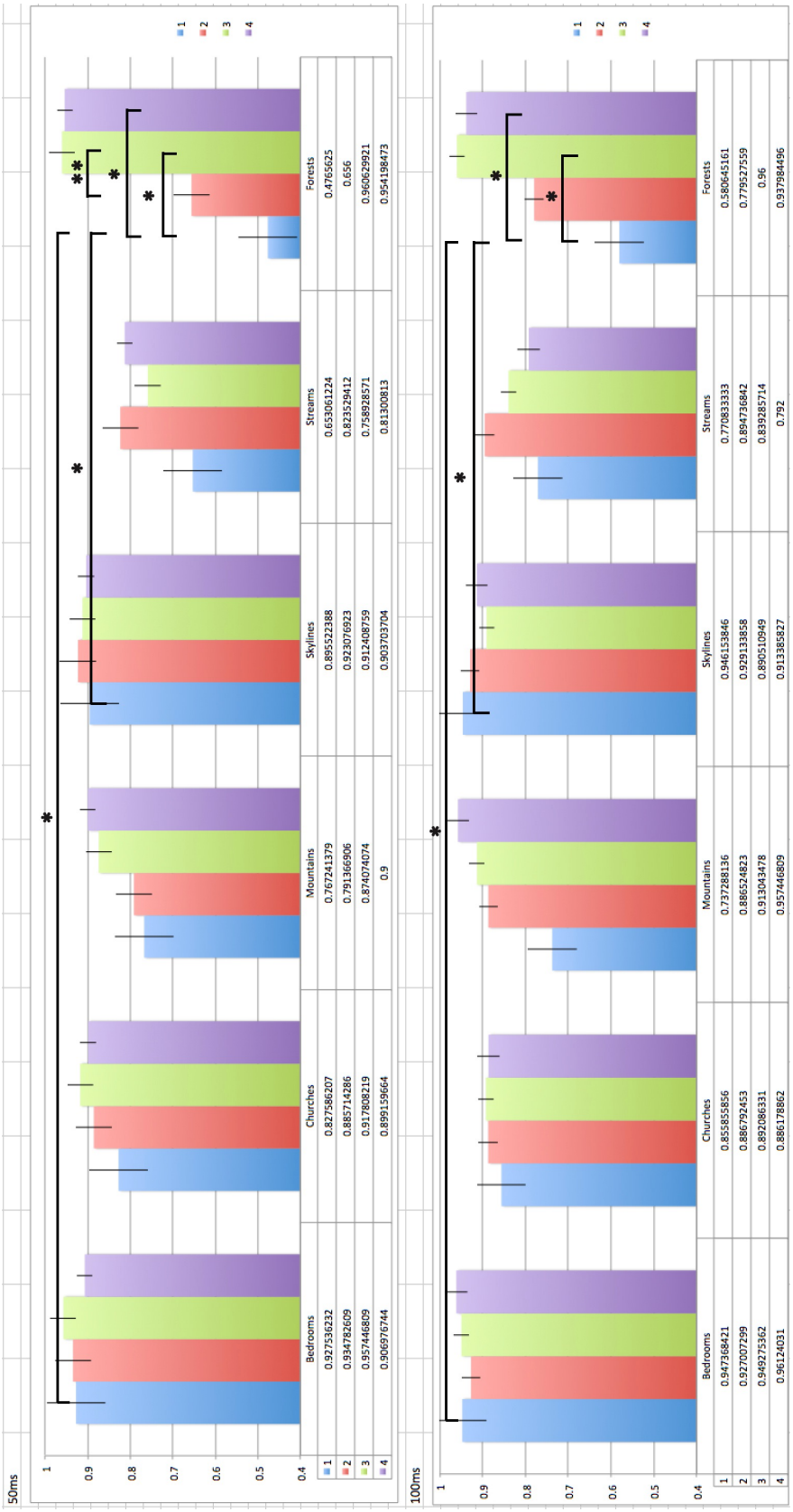

Figure 4f: Mean accuracy for the six categories, four SF, and two durations, where chance = 17%. Single asterisks indicate a significant difference as defined by the Tukey's value of .310 at an alpha of .05, and double asterisks indicate marginal significance.

A Tukey's Post Hoc test was performed on pairwise comparisons from the significant three-way interaction between category, SF, and duration in the accuracy ANOVA. This test yielded a Tukey's value of .310 at an alpha of .05, revealing that at 50ms, mean accuracies for SF 1 and SF 4 forests were significantly different from one another, as were SF 1 and SF 3 forests, and the difference between SF 2 and SF 3 forests was marginally significant (.305 < .310). At 100ms, mean accuracies for SF 1 and SF 4 forests were significantly different, as were SF 1 and SF 3 forests. At both 50ms and 100ms, mean accuracies for SF 1 forests and SF 1 bedrooms were significantly different, as were SF 1 forests and SF 1 skylines (see figure 4f). Taken together, these post-hoc tests reveal that differences within the scene category of forests, particularly LSF-filtered forests, largely contribute to the significant three-way interaction for accuracy.

In order to find the source of this the three-way interaction, additional analyses of variance were performed. Separate 4 x 2 repeated measures ANOVA, in which SF and duration were the factors were conducted on mean accuracy at each of the four SF conditions and the two durations for each of the categories individually. For churches and skylines, there were no significant main effect or interactions between the factors. For bedrooms, there was just a significant main effect of duration ($F(1, 1) = 4.783$, $p = .038$). For mountains, there was was a significant main effect of SF ($F(1, 3) = 6.406$, p .001) and a significant interaction between SF and duration ($F(1, 3) = 4.559$, $p = .005$). For streams, there were significant main effects of SF ($F(1, 3) = 5.596$, $p = .002$) and of duration ($F(1, 1) = 5.915$, $p = .023$), but no significant interaction. For forests, there were significant main effects of SF ($F(1, 3) = 34.183$, $p < 0.001$) and of duration ($F(1, 1) = 7.805$, $p = 0.009$), as well as a significant two-way interaction between SF and duration ($F(1, 3) = 3.605$, $p = .017$).

One last analysis, a 2 x 4 x 2 repeated measures ANOVA, was performed to look at differences between the natural and manmade categories — which were defined *a priori* — the four SF, and the two durations. There was a significant main effect of manmade/natural ($F(1, 1) = 45.720$, $p < .001$). There was a significant two-way interaction between manmade/natural and SF ($F(1, 3) = 24.628$, $p < .001$) and a marginally significant interaction between manmade/natural and duration ($F(1, 1) = 3.505$, $p = .076$) (see figures 5a, 5b, & 5c).

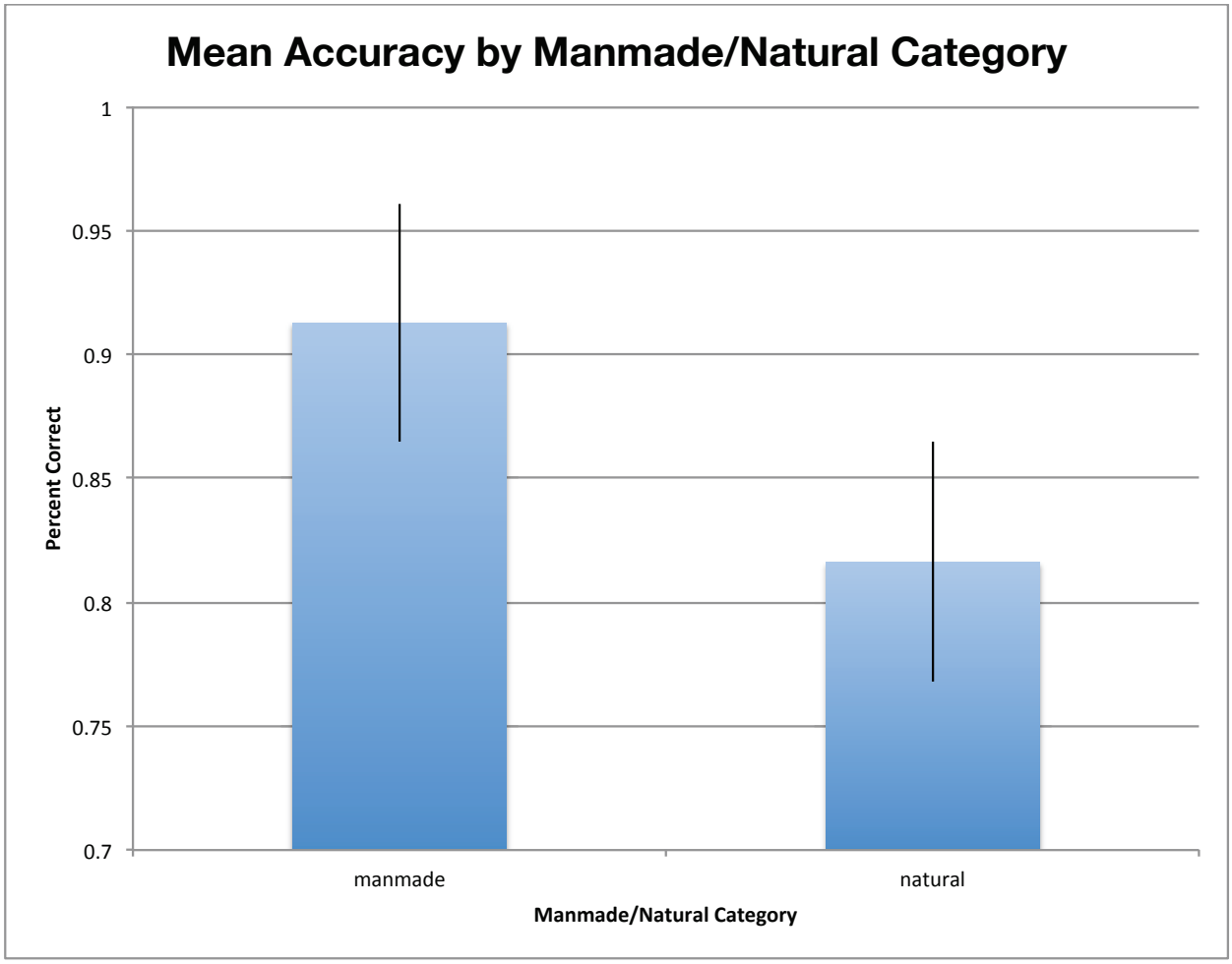

Figure 5a: Mean accuracy for the manmade (bedrooms, churches, skylines) versus natural (mountains, streams, forests) categories. Chance = 17%.

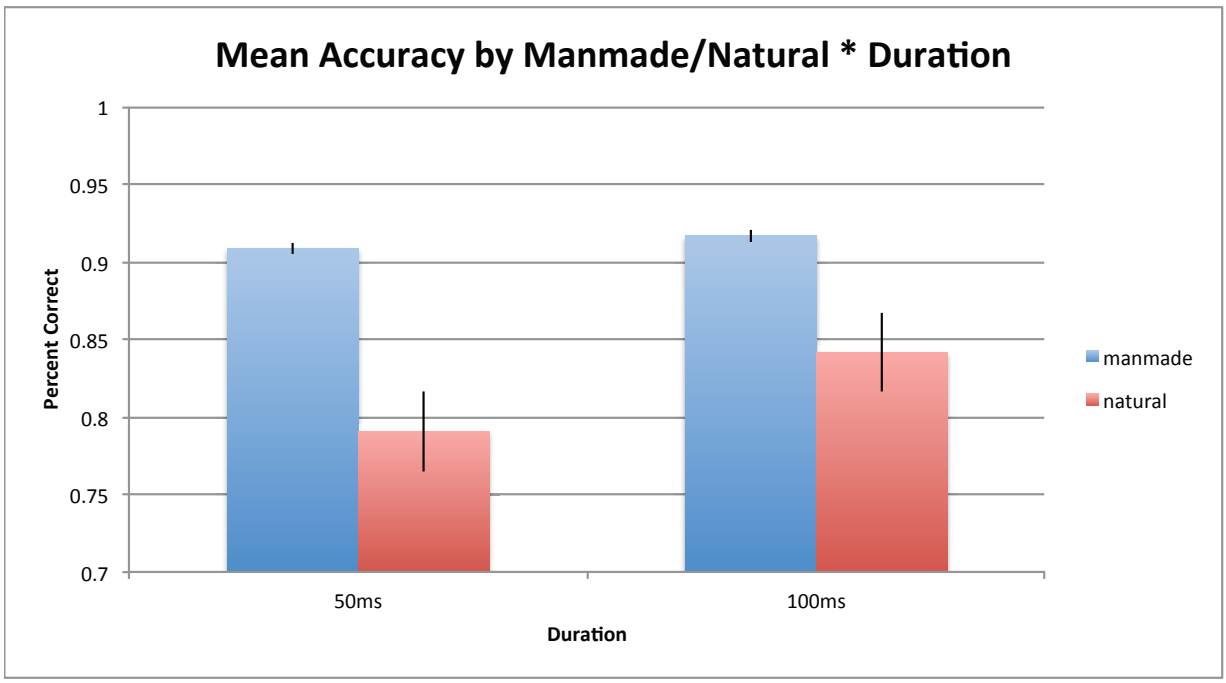

Figure 5b: Mean accuracy for the manmade and natural categories at each duration. Note that the two-way interaction is only marginally significant. Chance = 17%.

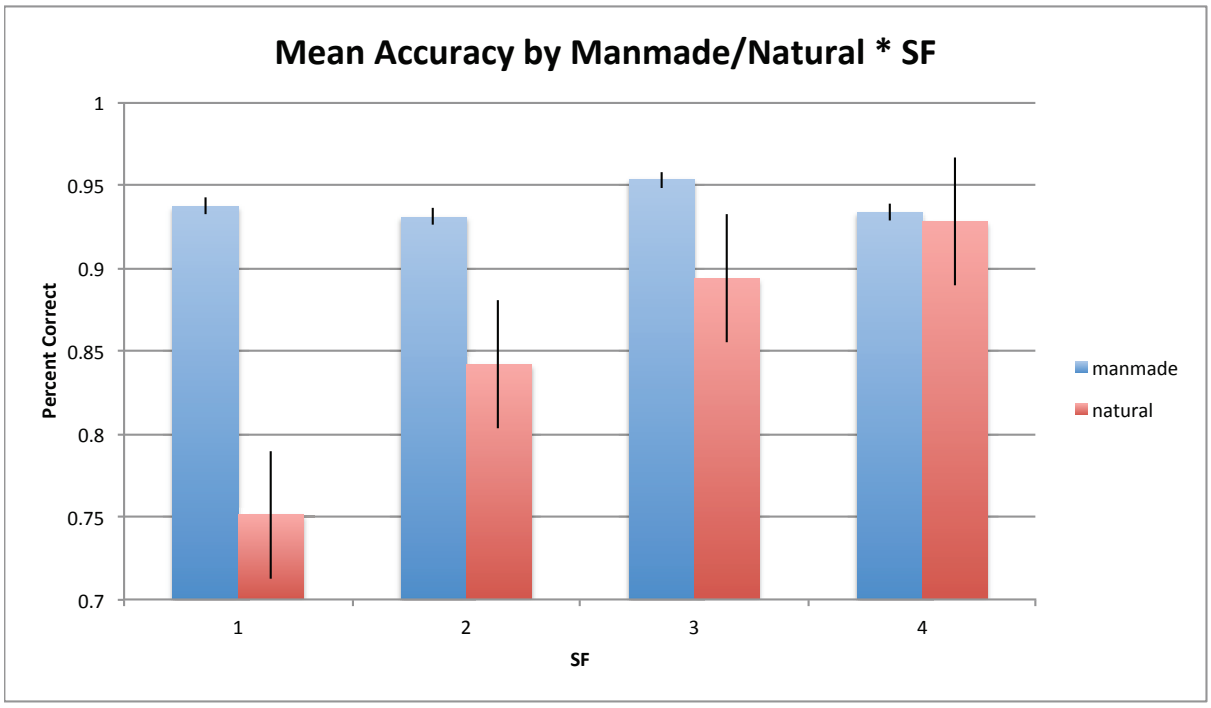

Figure 5c: Mean accuracy for the manmade and natural categories at each SF. The two-way interaction is significant. Chance = 17%.

Discussion

This visual psychophysics experiment sought to assess the role of spatial frequency in rapid scene processing. In a difficult six-way scene categorization task, we examined the effect of LSF and HSF filtering on subjects' performance at a very short presentation duration (50ms) and a slightly longer one (100ms). Performance was measured as reaction time and accuracy. Prior to the task, subjects were trained extensively on both intact and SF-filtered images of scenes from six categories, which were selected to span a large range of spatial frequencies inherent to these scene types and to represent variation in the holistic 'spatial envelope' (Oliva & Torralba 2001). Therefore, our scene stimuli were selected and parameterized to include categories that were natural and manmade, close and far, and closed and open, as Kravitz, Peng, and Baker (2011) did.

We predicted that at the shorter 50ms presentation duration, there would an inverse linear relationship between subjects' performance and SF, such that across the four SF conditions, LSF scenes would elicit the best performance and HSF scenes the worst. At the longer duration, we predicted that the facilitation in performance for LSF scenes would disappear, such that performance would be equivalent across the four SF conditions. These predictions were based on a number of studies that have shown that an LSF-sensitive top-down mechanism facilitates performance in rapid visual recognition, and in particular, that LSF are sufficient for rapid scene processing. Our prediction of facilitated performance for LSF over HSF scenes at the fast but not slower duration was also based on the results of Bar et al. (2006) that showed that LSF

components are processed in advance of HSF components of scenes during rapid scene recognition.

We found that there were no significant two-way or three-way interactions between category, SF, and duration for RT, which motivated us to take accuracy as our primary measure of subjects' performance. The first main finding in this study was that there was not a significant interaction between SF and duration for accuracy across all six categories. That is, performance facilitated by the presence of a larger proportion of low or high spatial frequencies in the scenes did not differ as a function of the duration that the images were displayed for. There are a number of methodological reasons why this could be the case. It is possible that 50ms was an adequately short image presentation duration to assess rapid scene processing, but that 100ms was not adequately long, and that at 100ms scene processing was subserved by the same underlying mechanisms as at 50ms. In fact, our finding of a significant main effect of duration for accuracy across the categories, revealing better performance overall at 100ms than 50ms durations, indicates that that 100ms may have actually represented a sweet spot for rapid scene processing. However, we can not make any direct conclusions about how the SF components of the scenes contributed to this difference. In future experiments, it will be beneficial to introduce longer durations, such as 500ms, in order to dissociate between effects arising from SF-sensitive mechanisms subserving rapid scene processing and those from other non-rapid processing.

Taken together with the finding of a main effect of category, the finding of a significant three-way interaction between category, SF, and duration for accuracy suggests that within certain scene

types, there may be performance facilitated by a combination of SF and duration related factors. The Tukey's post-hoc test revealed that forests in particular demonstrated such an effect, such that subjects were most accurate for forests with a higher proportion of high spatial frequencies, and least accurate for forests with higher proportion of low spatial frequencies. This effect was present at both the 50ms and 100ms durations, further supporting the need for future investigations which use longer presentation durations. The post-hoc tests also showed that, at both durations, performance for the lowest SF filtered forests was significantly worse than that for the lowest SF filtered bedrooms, and worse than that for the lowest SF filtered mountains. The source of this effect is unknown, but it may have been that LSF forests were just particularly difficult for subjects to categorize in general. Indeed, forests showed the lowest accuracies of all the categories, as was revealed by the significant main effect of category. Furthermore, LSF filtered scenes showed the lowest accuracy and highest RT across all categories. Therefore, it may be that at the durations used in this study, LSF filtered scenes elicit poorer performance in general than HSF filtered scenes.

The significant main effect of category and category- and duration- specific differences in performance may be explained in part by the results of our last analyses, which separate the categories into manmade and natural scenes, in line with our a priori assumptions based on Kravitz, Peng & Baker (2011). Here, there is a main effect of natural/manmade such that performance is significantly better overall for manmade than natural scene categories. Although the two-way interaction between manmade/natural and duration is only marginally significant, we can speculate that performance on manmade scenes is consistent across the durations, whereas for natural scenes performance is better at 100ms than at 50ms. It may be possible that the mechanism subserving rapid scene processing struggles with natural scenes, in which

they are inherently a lower proportion of high spatial frequencies, but less so with manmade scenes, in which there are inherently a higher proportion of high spatial frequencies. That is, only if we are assuming that high spatial frequencies do in fact facilitate rapid scene processing.

Alternatively, the significant two-way interaction between manmade/natural and SF may support another explanation for overall better performance on manmade than natural scenes. For manmade scenes, performance was relatively constant across the four SF conditions, whereas for natural scenes performance was poorest for scenes with the highest proportion of LSF and best for scenes with the highest proportion of HSF, with a linear relationship between performance and SF. Furthermore, performance was almost equivalent for manmade and natural scenes at the highest SF condition. In other words, it appears that for natural scenes, but not manmade scenes, HSFs facilitate rapid scene processing and LSFs do not. Conversely, for manmade scenes, both LSF and HSF are sufficient for rapid scene processing. This natural/ manmade split between the categories may account for our previous findings of better performance overall for HSF-filtered scenes. It is also worth considering that there was an unforeseen interaction between the predominately lower SFs inherent to natural scenes and the LSF filtering used in this study, that resulted in poorer performance.

Going forward, in future psychophysical and neuroimaging studies, it will be interesting to investigate how high spatial frequencies contribute to rapid scene processing. And, it will also be important to see how the type of scene — whether natural or manmade, or other parameterizations — and its inherent spatial frequency content and ‘spatial envelope’, contributes to this effect. For the moment, it is unclear why the findings of the current study do

not replicate others' findings of a LSF-dependent mechanism mediating rapid scene processing, and of LSF being processed prior to HSF in scenes. The finding of facilitated performance for HSF scenes in the 100ms condition may support the finding of Rajimehr et al. (2011) that the parahippocampal place area (PPA) responds preferentially to the high spatial frequencies in scenes. Further, Walther et al. (2011) were able to show that human subjects were able to recognize and categorize line drawings of scenes, demonstrating that the structure of scenes, devoid of *any* low spatial frequencies, was able to convey information on the probable semantic category (Walther et al. 2011).

Citations

Bar, M. & Aminoff, E. (2003). Cortical analysis of visual context. *Neuron*, 38, 347-358.

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

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

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., A., Schmid, M., Dale, A. M., Hämäläinen, M., Marinkovic, S., K., Schacter, D. L., Rosen, B. R., and Halgren, E. (2006). Top-down facilitation of visual recognition. *PNAS* 103 (2) 449-454

Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends in Cognitive Sciences* 11: 280–289.

Bullier, J. (2001). Feedback connections and conscious vision. *Trends in Cognitive Sciences* 5 (9):369-370.

Egeth, H.E. & Santee, J.L. (1982). Do Reaction Time and Accuracy Measure the Same Aspects of Letter Recognition? *Journal of Experimental Psychology*. 8:489-501

Ganaden, R.E., Mullin, C.R. and Steeves, J.K.E. (2013). Transcranial Magnetic Stimulation to the Transverse Occipital Sulcus Affects Scene but Not Object Processing. *Journal of Cognitive Neuroscience,* 25, 6, 961-968.

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

Kravitz, D.J., Peng C.S., Baker, C.I. (2011). Real-world scene representations in high-level visual cortex – it's the spaces not the places. *J. Neurosci.* 31, 7322–7333

Kveraga K, Boshyan J, Bar M. (2007). Magnocellular projections as the trigger of top-down facilitation in recognition. *Journal of Cognitive Neuroscience* 27: 13232–13240.

Mullin, C. R., & Steeves, J. K. E. (2011). TMS to the lateral occipital cortex disrupts object processing but facilitates scene processing. *Journal of Cognitive Neuroscience*, 23, 4174–4184.

Oliva, A., & Torralba, A. (2001). Modeling the Shape of the Scene: a Holistic Representation of the Spatial Envelope. International Journal in Computer Vision, 42, 145-175.

Oliva, A. & Torralba, A. (2006). Building the gist of a scene: the role of global image features in recognition. Chapter in *Progress in Brain Research.* Vol 155. 23-36.

Oliva, A. (2013). Scene Perception. Chapter in the New Visual Neurosciences, Eds John S. Werner and Leo. M. Chalupa. *MIT Press*.

Park, S. J., Konkle, T. & Oliva, A. (2014). Parametric Coding of the Size and Clutter of Natural Scenes in the Human Brain. *Cerebral Cortex*.

Peyrin, C, Chokron, S, Guyader, N, Gout, O, Moret, J, Marendaz, C. (2006). A neural correlates of spatial frequency processing: a neuropsychological approach. *Brain Res* 1073–1074: pp. 1-10

Peyrin, C, Michel, CM, Schwartz, S, Thut, G, Seghier, M, Landis, T, Marendaz, C, Vuilleumier, P. (2010). The neural substrates and timing of top-down processes during coarse-to-fine categorization of visual scenes: a combined fMRI and ERP study. *J Cogn Neurosci* 22: pp. 2768-2780

Potter, M. C., & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology*, 81, 10–15.

Rajimehr, R., Devaney, K.J., Bilenko, N.Y., Young, J.C., Tootell, R.B. (2011). The "parahippocampal place area" responds preferentially to high spatial frequencies in humans and monkeys. *PLoS Biol* 9:e1000608. CrossRef Medline.

Schyns, P.G., Oliva, A., 1994. Evidence for time- and spatial-scale-dependent scene recognition. *Psychol. Sci.* 5, 195–201.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381(6582), 520–522.

Vuilleumier, P., Armony, J.L., Driver, J., Dolan, R.J. (2003). Distinct spatial frequency sensitivities for processing faces and emotional expressions. *Nature neuroscience* 6 (6), 624-631.

Walther, D.B., Chai, B., Caddigan, E., Beck, D.M., and Fei-Fei, L. (2011). Simple line drawings suffice for functional MRI decoding of natural scene categories, *PNAS* 108 (23): 9661-9666.

Woodhead, Z. V. J., Wise, R. J. S, Sereno, M., & Leech, R. (2011). Dissociation of Sensitivity to Spatial Frequency in Word and Face Preferential Areas of the Fusiform Gyrus. *Cerebral Cortex.* 21(10):2307-12.