# Temporal Processing of Spatial Frequencies in Visual Word, Object, and Place Recognition

**Clémence Bertrand Pilon (clemence.bertrand.pilon@umontreal.ca)**

Department of Psychology
Université de Montréal, Montreal, Quebec, Canada
Centre de recherche de l'Institut universitaire de gériatrie de Montréal
Montreal, Quebec, Canada,
Centre interdisciplinaire de recherche sur le cerveau et l'apprentissage (CIRCA), Department of Psychology,
Université de Montréal, Montreal, Quebec, Canada

**Gabriela Milanova (gabriela.milanova@umontreal.ca)**

Department of Psychology
Université de Montréal, Montreal, Quebec, Canada
Centre de recherche de l'Institut universitaire de gériatrie de Montréal
Montreal, Quebec, Canada,
Centre interdisciplinaire de recherche sur le cerveau et l'apprentissage (CIRCA), Department of Psychology,
Université de Montréal, Montreal, Quebec, Canada

**Anjali Singh (anjali.singh@umontreal.ca)**

Department of Psychology
Université de Montréal, Montreal, Quebec, Canada
Centre de recherche de l'Institut universitaire de gériatrie de Montréal
Montreal, Quebec, Canada,
Centre interdisciplinaire de recherche sur le cerveau et l'apprentissage (CIRCA), Department of Psychology,
Université de Montréal, Montreal, Quebec, Canada

**Martin Arguin (martin.arguin@umontreal.ca)**

Department of Psychology
Université de Montréal, Montreal, Quebec, Canada
Centre de recherche de l'Institut universitaire de gériatrie de Montréal
Montreal, Quebec, Canada,
Centre interdisciplinaire de recherche sur le cerveau et l'apprentissage (CIRCA), Department of Psychology,
Université de Montréal, Montreal, Quebec, Canada

## Abstract

**The temporal dynamics of spatial frequency (SF) processing in visual recognition were studied through three studies involving word (Exp.1), object (Exp.2) and place (Exp.3) recognition tasks in normal adult observers. They had to report the target stimulus presented in a 200 ms display using a four-alternative-forced-choice task (4AFC). The stimuli were made of an additive combination of the signal (target stimuli) and of a visual white noise patch wherein the signal-to-noise ratio (SNR) varied randomly across stimulus duration. Four SF conditions were defined with center frequencies of 1.2, 2.4, 4.8 and 9.6 cycles per degree. Contrary to the "coarse-to-fine" theory (Bar, 2003), the results indicated a more complex, non-sequential processing pattern which varies across stimulus classes. In Exp.1 (words), the highest SF range dominates early processing, with a shift toward lower SFs later on. In Exp.2 (objects), initial processing was dominated by the 4.8 cpd band, followed briefly by 9.6 cpd, then 1.2 cpd, and finally 2.4 cpd. In Exp.3 (places), processing was dominated first by the 2.4 cpd band, followed by 9.6 cpd and 4.8 cpd bands. These findings challenge the coarse-to-fine model. Also, all studies found that SF processing is modulated not only by the passage of time but also by the oscillatory frequency spectrum of the stimulus, indicating that visual recognition involves an intricate interaction of temporal and SF factors.**



**Keywords:** Visual recognition; Spatial frequency processing


## Introduction

Over the past decades, numerous studies have focused on understanding the neural mechanisms involved in visual recognition, with particular emphasis on recognizing words, objects, and places. To fully comprehend how the human visual system processes and identifies such diverse stimuli, it is crucial to examine the role of spatial frequencies (SFs; i.e the periodic distributions of light and dark in visual stimuli), which play a key role in visual processing. According to the coarse-to-fine theory (Bar, 2003; Caplette et al., 2017; Kauffmann et al., 2014), low SFs are primarily involved in the rapid, coarse processing of visual stimuli. This early processing is thought to provide a quick “initial guess” about the identity of the stimulus, enabling its broad categorization. High SFs would be responsible for more detailed but slower processing, allowing the specific identification of the stimulus. Since coarse-to-fine processing is widely thought to play a critical role in visual recognition, a detailed assessment of how SFs are processed over time in visual recognition is essential. For this purpose, the present study uses the random temporal sampling method (Arguin et al., 2021) with spatially filtered stimuli in tasks of word, object, and place recognition, each of which being known to involve distinct high-level brain areas in humans (Grill-Spector and Malach, 2004).

## Methods

Forty-eight adults (16 per experiment) with normal or corrected vision were recruited. They were asked to recognize the target (words, objects, or places) presented for a duration of 200 ms among four alternatives displayed afterward. Targets were filtered according to one of four SF conditions using a bandpass Butterworth filter. The center frequencies and cutoffs (in parentheses for each condition) were 1.2 (0.9–1.5), 2.4 (1.8–3.0), 4.8 (3.6–6.0), and 9.6 (7.2–12) cycles per degree. Response alternatives were broadband images. The target images were a linear summation of the signal (spatially filtered target) and a white noise field. The signal-to-noise ratio (SNR) varied according to a random sampling function made by the integration of sine waves with frequencies ranging between 5 and 55 Hz in steps of 5 Hz with random amplitudes and phases. Thus, the visibility of the target among the noise varied randomly throughout its exposure duration. New, independent SNR functions and white noise fields were generated for each trial.

## Results

Each graph of Fig. 1 shows the processing efficiency through time for each SF range and each stimulus class. For words (Fig. 1 a) the initial greatest level of efficiency was for the

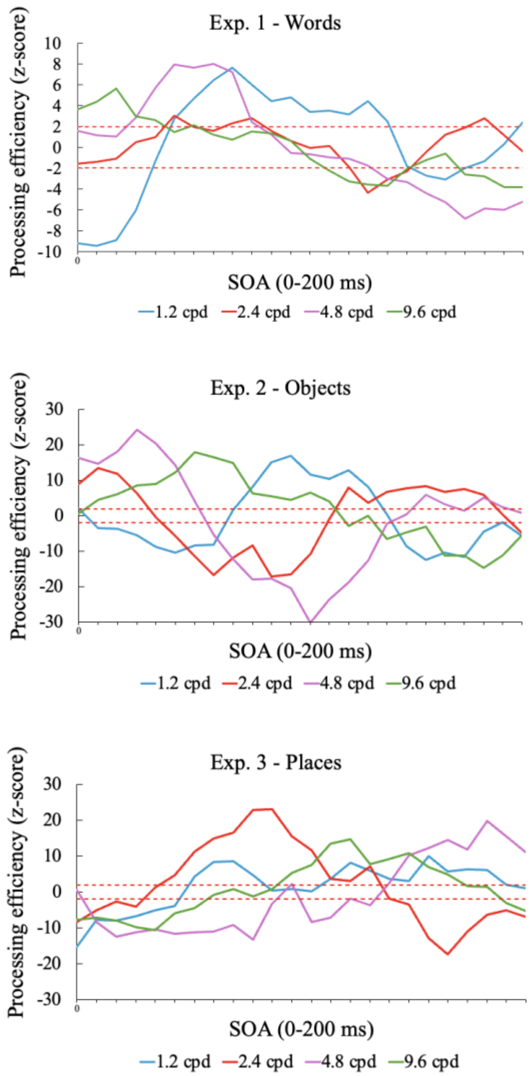


**Figure 1.** Classification images (CIs) of processing efficiency as a function of stimulus onset asynchrony (SOA) for each SF condition (condition 1 – 1.2 cpd, condition 2 – 2.4 cpd, condition 3 – 4.8 cpd, and condition 4 – 9.6 cpd) for words (a), objects (b) and places (c).

condition with 9.6 cpd center frequency (highest SF range), followed 4.8 cpd, each for a 33 ms duration. Only then, starting at 67 ms does condition with 1.2 cpd center frequency (lowest SF range) lead to the greatest efficiency until 150 ms, which is then followed by 2.4 cpd and then, by a brief reprieve of 1.2 cpd on the last display frame. This pattern is almost the opposite of what the «coarse-to-fine» hypothesis suggests. For objects the results show (Fig. 1b) that the 4.8 cpd center frequency condition is the most effective from target onset up to 50 ms, followed by 9.6 cpd for a duration of 42 ms, then 1.2 cpd for 60 ms, and ending with 2.4 cpd lasting 54 ms. For places (Fig. 1c), the earliest efficiency significantly above 0 is at 50 ms, where 2.4 cpd center frequency dominates up to 108 ms. This is followed by the 9.6 cpd condition from 117 to150 ms, and then 4.8 cpd until 200 ms. The data patterns for objects and places indicate non-sequential processing orders for the different SF bands and both are also inconsistent with the notion of coarse-to-fine processing. For each experiment, time- frequency classification images were also calculated, which show that SF processing effectiveness varies not only according to time and stimulus class, but also according to the temporal frequency spectrum of SNR oscillations. The Fourier spectra of the time-frequency classification images of individual participants were presented to a support-vector-machine (SVM) algorithm (with leave-one-out cross validation) which had to identify the SF condition that produced the data pattern it was exposed to. Data classification accuracy for Exp.1 reached 90.6% (vs. 25% chance performance) using only 8.0% of the available features. For Exp.2, the classifier reached the same response accuracy, while using 8.6% of the available features. For Exp.3, the maximum classification performance achieved was 90.6% using 7.8% of the features available.

## Discussion

The results with all stimulus classes point to a processing order incompatible with the notion of coarse-to-fine processing. Our findings also show that the temporal frequency spectrum of target visibility has a major impact on SF processing and that it interacts with the passage of time. By using an SVM learning algorithm, we were able to successfully discriminate among the SF conditions based on their characteristic temporal features. This means that in recognizing those stimuli, the processing of different SF ranges is subserved by dissociable oscillatory neural mechanisms.

## Acknowledgements

Supported by grants from the Fonds de Recherche Québec—Nature et Technologie and the Natural Sciences and Engineering Research Council of Canada (NSERC) to Martin Arguin.

## References

Arguin, M., Ferrandez, R., & Massé, J. (2021). Oscillatory visual mechanisms revealed by random temporal sampling. *Scientific Reports*, *11*(1), 1-14. https://doi.org/10.1038/s41598-021-00685-w

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience*, *15*(4), 600-609. https://doi.org/10.1162/089892903321662976

Caplette, L., Wicker, B., Gosselin, F., & West, G. L. (2017). Hand position alters vision by modulating the time course of spatial frequency use. *Journal of Experimental Psychology: General*, *146*(7), 917-923. http://dx.doi.org/10.1037/xge0000296

Grill-Spector, K. & Malach, R. (2004). The human visual cortex. *Annual Review of Neuroscience, 27*, 649–677.

Kauffmann, L., Ramanoël, S., & Peyrin, C. (2014). The neural bases of spatial frequency processing during scene perception. *Frontiers in Integrative Neuroscience*, 8, 37. https://doi.org/10.3389/fnint.2014.00037