

Modeling the Visual Processing System with Gabor Filters and Spiking Neurons

Jaeho Cho

MA391 - Theoretical Neuroscience

The Cooper Union for the Advancement of Science and Art

New York City, NY

jaeho.cho@cooper.edu

Abstract—This paper presents a computational model of early visual processing using Gabor filters and spiking neurons to simulate encoding and decoding mechanisms in the visual system. Gabor functions model the receptive fields of V1 neurons, and spike trains are generated via a Poisson process with refractory periods. Visual stimuli are encoded by populations of neurons with varying orientations and spatial frequencies, and reconstructed linearly. Results demonstrate the model's ability to preserve key image features, with reconstruction quality improving as population size increases. The model offers insights into fundamental neural dynamics.

Index Terms—Theoretical Neuroscience, Spiking Neuron, Gabor Filter

I. INTRODUCTION

Encoding refers to the process by which external visual stimuli are represented as patterns of neural activity. This begins when photoreceptors in the retina convert light into electrical signals. These signals are then processed through various neural pathways, including the lateral geniculate nucleus (LGN) and the primary visual cortex (V1), where features such as orientation, spatial frequency, and color are extracted. The efficient coding hypothesis suggests that the visual system optimizes these representations to maximize information transmission while minimizing redundancy, effectively compressing the vast amount of visual data encountered in natural environments [1].

Decoding involves interpreting these neural representations to reconstruct or predict the original visual stimuli. Techniques such as functional magnetic resonance imaging (fMRI) have been used to decode visual experiences by mapping specific neural activation patterns to particular visual inputs. For instance, studies have demonstrated the ability to reconstruct images by decoding spike trains in a population of Retinal Ganglion Cells (RGC) as well as from fMRI data [2], [3].

This paper presents a computational approach to modeling the early visual processing system, focusing on the encoding and decoding of visual information. The receptive fields of neurons in the primary visual cortex (V1) are sensitive to features like orientation, spatial frequency, and phase which can be effectively modeled using Gabor functions. The spiking neuron models replicate the generation of action potentials in real neurons, providing insights into the dynamics of neural coding and information processing [4].

II. METHODS

A grayscale image from Unsplash was used as the primary stimulus for this paper as it includes natural scenes with various textures and structures [5]. A binary sinusoidal grating was generated as a secondary stimulus, which serves as a reference for evaluating the Gabor filter responses.

The spatial receptive fields modelled as Gabor functions are of the form:

$$D_s(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2} \cos(kx - \varphi)\right) \quad (1)$$

where σ_x and σ_y define the spatial extent, k is the preferred spatial frequency, and φ is the preferred phase. In this paper, the phase φ is set to $\frac{\pi}{2}$ to create a filter that splits the ON and OFF regions of the receptive field to each side of the center. The general appearance of the Gabor function is shown in Fig. 1.

The firing rate for a single neuron is estimated by computing the linear response:

$$r_{\text{est}} = \int dx dy D(x, y) s(x, y) \quad (2)$$

where $D(x, y)$ is the Gabor function and $s(x, y)$ is the stimulus. The response generated by this estimation is demonstrated in Fig. 2; and the relative size of this gabor filter to the natural image is shown in Fig. 3.

Each neuron in the model generates a spike train based on a scaled estimated firing rate (r_{est}); the spike trains are generated using a homogeneous Poisson process with an absolute refractory period of 1 ms and relative refractory period with mean 10 ms [6]. The firing rate is then averaged over a time window to obtain the more realistic estimated firing rate.

Two populations of neurons are used to model the visual system: a population of gabor filters with an orientation of $\varphi = \pi$ rad and another with $\varphi = -\frac{\pi}{2}$; these orientations differentiated the preferred orientations of the filters, acting as vertical and horizontal edge detectors. The population response is given in the form of a 2D array appearing as the edges of the image, as shown in Fig. 4. The figure illustrates the necessity of using populations of different orientations to capture the full range of features in the image as the edges of orthogonal to the preferred orientation of the filter are not detected.

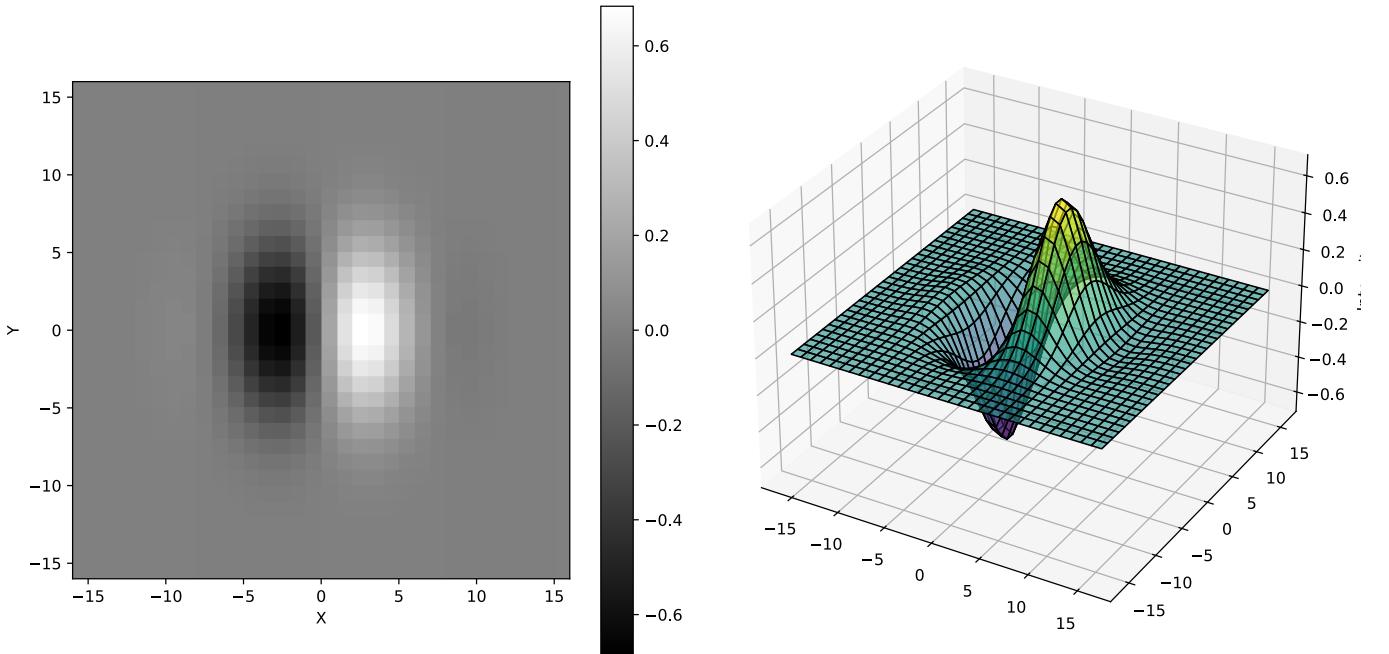


Fig. 1: The left side illustrates the gabor filter as a colormap with brighter values corresponding to ON regions and darker values corresponding to OFF regions. The right side shows the same gabor filter in a 3D representation.

The primary reconstruction method used in this paper is a simple cumulative sum in the direction of the populations preferred orientation, where the average estimated firing rates of the neurons are used to reconstruct the original image; these reconstructed images from each population are then averaged to create the final decoded image.

III. RESULTS

Fig. 5 shows the various firing rates of neurons down the center of the image depending on the spike train generation model. As expected, the models with no refractory effects are more closely aligned with the original r_{est} values. The models with refractory effects appear to be a scaled and shifted version of the original r_{est} values. As the refractory effects cause the removal of spikes, the estimate firing rates should be lower at higher frequencies, and given the large relative refractory period, the estimated firing rate at even lower frequencies is also lower. The figure also illustrates how a longer period of decoding results in less deviation from the original r_{est} values.

The effect of changing the number of neurons in the population is illustrated in Fig. 6. For the first two reconstructions, the majority of the major features of the image are preserved, like the tree and edge of the building, however in the second reconstruction, the lamp post is not clearly distinguished. The relative entropy of the decoded images relative to the original image was computed to be 0.22, 0.20, and infinity for the first, second, and third images respectively.

In order for the image to be reconstructed from the population responses, they need to be scaled such that mean of the population response is 0. This is because the cumulative sum used to decode the response is a linear operation, and if the

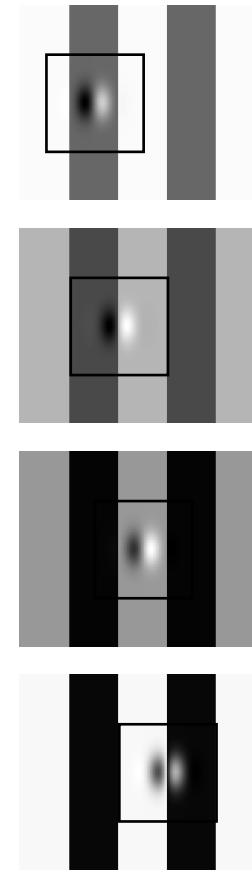


Fig. 2: The gabor overlayed on vertical gratings at different locations of the image. The first instance with the gabor filter entirely on the black strip outputs a response of 0; the second location with the filter split between a black strip under the OFF region and a white strip under the ON region outputs a response of approximately 115; The third and fourth instances output responses of 0 and -115, respectively.

mean of the population response is not 0, the reconstructed image will be biased towards the mean of the population response, causing a gradient effect. Histograms of the population responses are shown in Fig. 7. The histograms generated from the nonideal spike train generation models display a discrete distribution of values, this is understandable as the firing rates are derived from counting the number of spikes in



Fig. 3: Overlay of Gabor filter on natural scene. The black box outlines the total size of the neuron, while the black and white regions represent the ON and OFF responses of the filter.

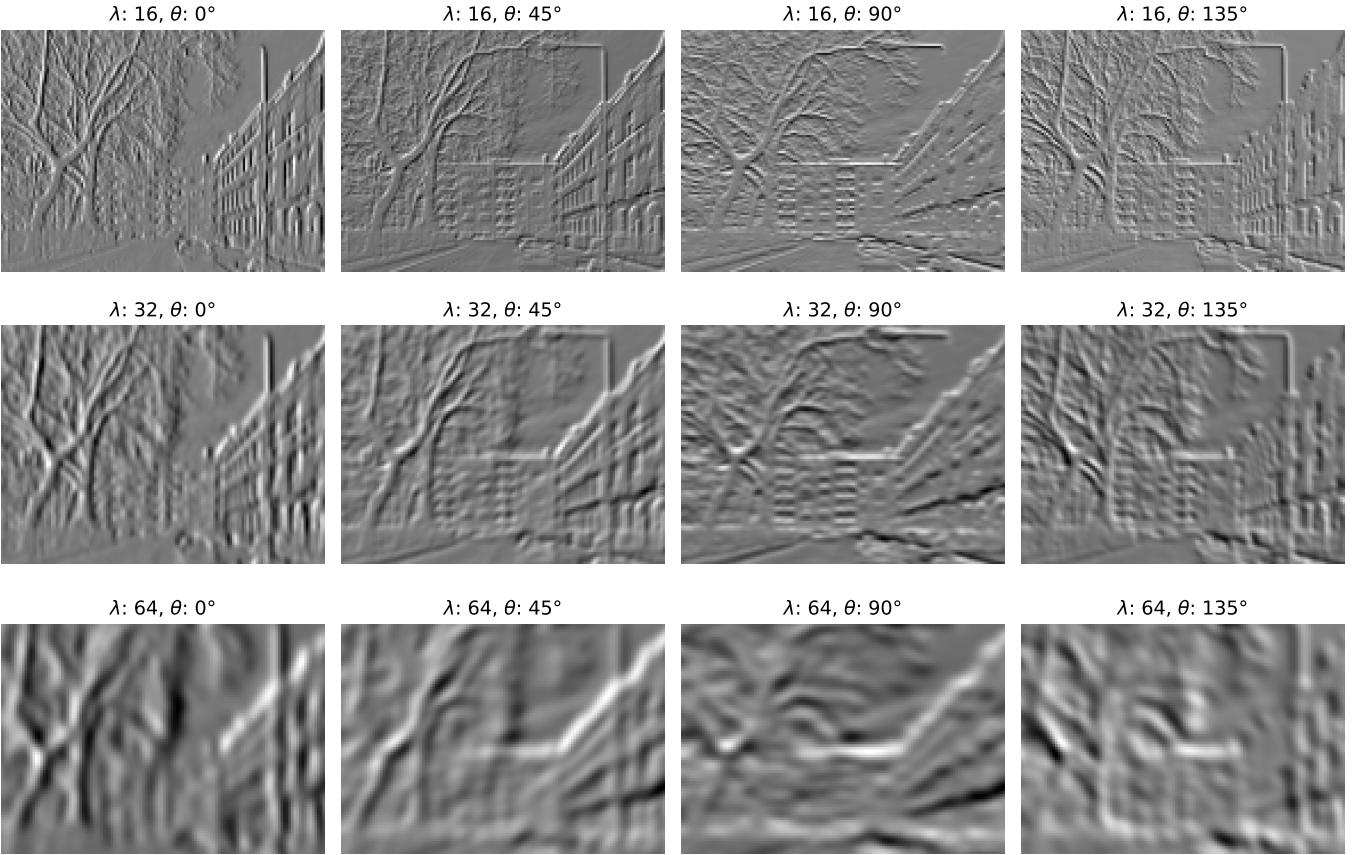


Fig. 4: Images encoded by a Gabor Filter Bank with wavelengths varying incrementally from 16 to 64 pixels and orientations varying from 0 to 135 degrees.

a time window, which vary by integers. The relative entropy was computer to be about 0.36 bits for the base population response.

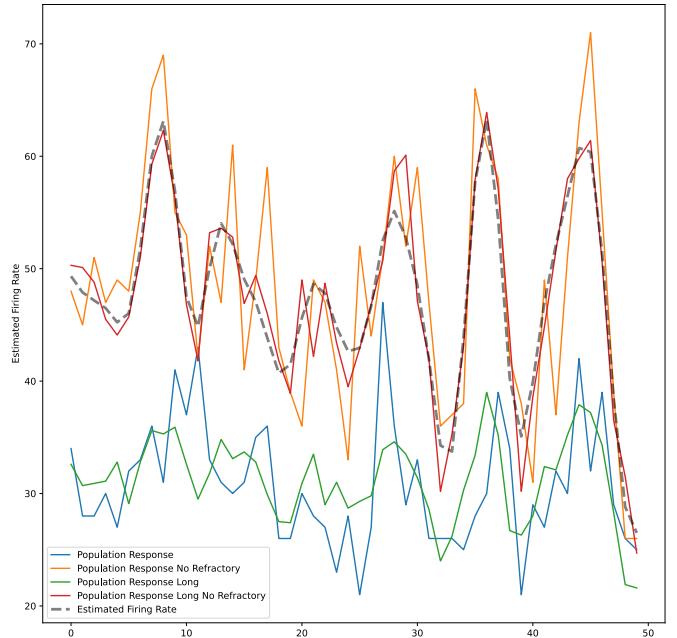


Fig. 5: Responses of neurons down the center of the image starting from the top to the 50th neuron.

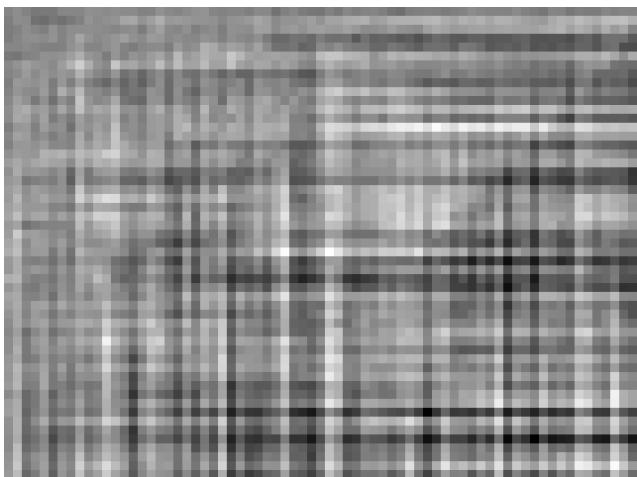
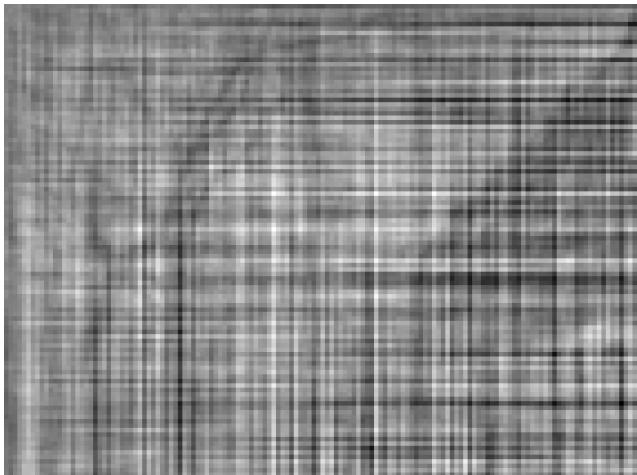
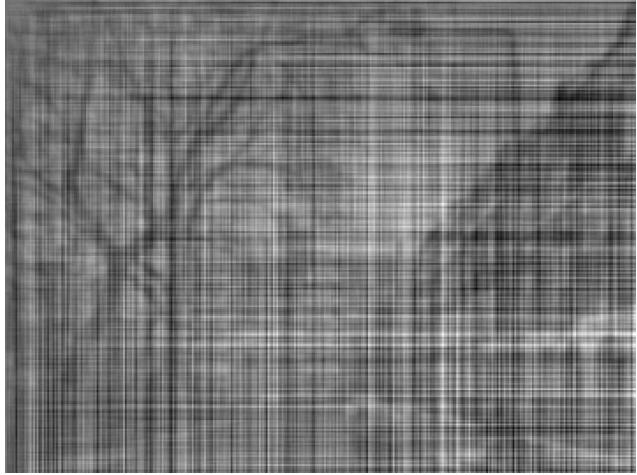


Fig. 6: The top image shows the decoded image using a population of 286 vertical neurons and 376 horizontal neurons (286×376). The image in the middle was decoded from an image of (114×150) neurons. The bottom image was decoded from (57×75) neurons.

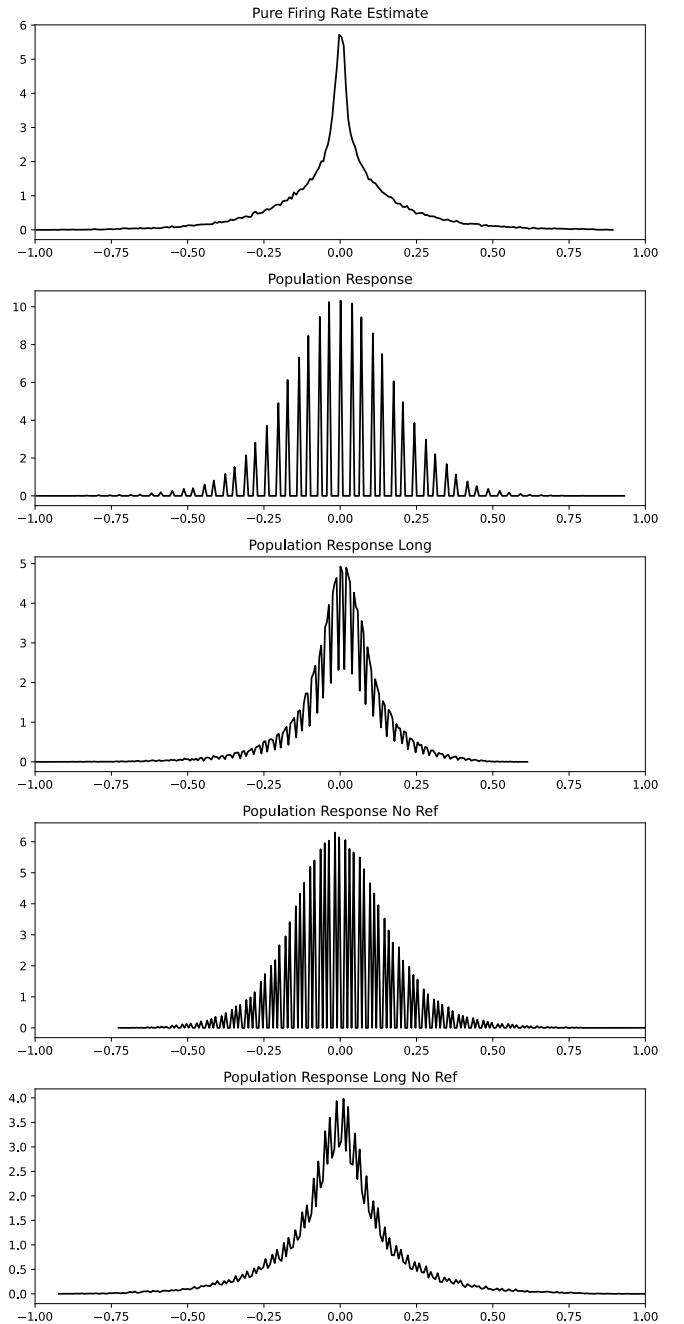


Fig. 7: Population responses of the neurons in the population. The top histogram shows the pure firing rate estimation without spike train generation. The other histograms show the distribution of firing rates generated from various spike train generation models. The line histograms are generated from 256 samples of the population response.

IV. DISCUSSION

One of the major causes of the imperfect reconstruction is the spike train generation adding noise to the population response. This noise in the population response results in streaks in the decoded images as an offset neuron causes a shift down the line of the cumulative sum. Fig. 8 shows how the population response is improved with longer periods of spike detection and absence of refractory effects.

Given the relatively small number of neurons in the population, it is surprising to see that the majority of the features of the image are preserved. As was illustrated in Fig. 6, the

increase in the number of neurons significantly improves the quality of the image and increases the signal to noise ratio.

This project integrates well-established principles of visual processing with computational methods where the use of Gabor functions as receptive field models provides a physiologically plausible basis for early visual processing. The experiments with multiple Gabor filters demonstrate how different scales and orientations contribute unique information about image structure, an observation that aligns with findings in biological systems where multiple populations of neurons respond selectively to different stimulus features [4].

Future work could explore the implementation of color information as well as time-receptive fields to decode videos.

REFERENCES

- [1] L. K. Loh and M. Bartulovic, “Efficient Coding Hypothesis and an Introduction to Information Theory,” 2014.
- [2] Y. Miyawaki *et al.*, “Visual Image Reconstruction from Human Brain Activity Using a Combination of Multiscale Local Image Decoders,” *Neuron*, vol. 60, no. 5, pp. 915–929, Dec. 2008, doi: 10.1016/j.neuron.2008.11.004.
- [3] Y. Zhang *et al.*, “Reconstruction of Natural Visual Scenes from Neural Spikes with Deep Neural Networks.” Accessed: Apr. 04, 2025. [Online]. Available: <http://arxiv.org/abs/1904.13007>
- [4] P. Dayan and L. F. Abbott, *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. in Computational Neuroscience. Cambridge, Mass.: MIT Press, 2001.
- [5] Unsplash, “Photo by Hulki Okan Tabak on Unsplash.” Accessed: Apr. 04, 2025. [Online]. Available: <https://unsplash.com/photos/grayscale-photo-of-cars-on-road-near-buildings-eY9wnVZOIWg>
- [6] “Project 1 - Spike-Train Generation — Theoretical Neuroscience.” Accessed: Apr. 04, 2025. [Online]. Available: <https://jaehho.github.io/Theoretical-Neuroscience/project-1.html>

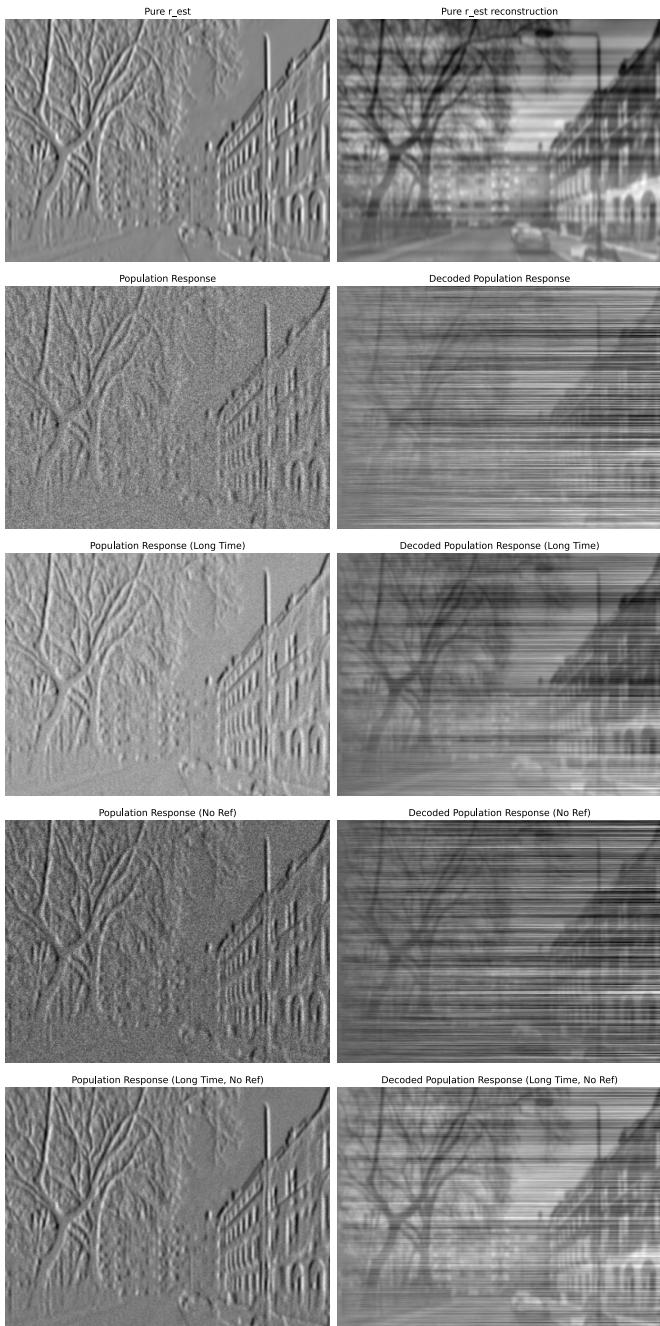


Fig. 8: Decoded Population Responses