A population receptive field modeling framework of sensory suppression in human visual cortex

Eline R Kupers (ekupers@stanford.edu)
Department of Psychology, 450 Jane Stanford Way
Stanford, CA 94305 USA

Insub Kim (insubkim@stanford.edu)
Department of Psychology, 450 Jane Stanford Way
Stanford, CA 94305 USA

Kalanit Grill-Spector (kalanit@stanford.edu)
Department of Psychology & Wu Tsai Neurosciences Institute, 450 Jane Stanford Way
Stanford, CA 94305 USA

Abstract:

When presenting multiple visual items simultaneously in the receptive field, the neurophysiological response is surprisingly lower than presenting the identical items sequentially. However, the underlying computations of this suppression effect are not well-understood. Here, we leveraged population receptive field (pRF) models to computationally test how linear, compressive spatial, or compressive spatiotemporal summation contributes to suppression at the voxel level. We collected two fMRI experiments in 10 subjects: (i) retinotopy to estimate pRFs and (ii) an experiment with simultaneous or sequential stimuli, both varying in duration and size. In V1, there was no simultaneous suppression and responses were larger for bigger stimuli. This was well-predicted by linear pRFs. However, the linear model failed to capture larger responses in V1 for brief vs long durations. In V2 and high-level areas, responses were lower for simultaneous vs sequential stimuli, larger for brief vs long durations, and did not increase much with size. Both compressive pRF models predicted simultaneous suppression and the effect of stimulus size, but only the compressive spatiotemporal model predicted the effect of duration. Our results suggest that compressive spatiotemporal pRFs are necessary to predict responses in visual cortex to simultaneous vs sequential stimuli and underscore the power of pRF models for providing new insights into spatiotemporal computations of sensory suppression.

Keywords: Visual cortex; sensory suppression; nonlinear summation; population receptive fields; fMRI.

Introduction

A robust, yet intriguing phenomenon in cognitive neuroscience is that responses in high-level visual areas are lower for simultaneously presented visual items than for the same items presented sequentially (Miller, Gochin, & Gross, 1993; Kastner, de Weerd, Desimone, & Ungerleider, 1998; Reynolds, Chelazzi, & Desimone, 1999; Kastner et al., 2001; N. Y. Kim, Pinsk, & Kastner, 2021). The prevailing hypothesis is that suppression reflects competition between multiple visual items in the receptive field (Desimone & Duncan, 1995). However, this competition hypothesis has not been computationally tested. Here, we operationalize competition using computational modeling of population receptive fields (pRFs) and test predictions with fMRI measurements at the single voxel level.

Methods

Participants. Ten volunteers (6 female, ages 22-53 years), who gave written informed consent, participated in two fMRI sessions.

Session 1: Retinotopic mapping. Observers viewed sweeping bar stimuli with cropped cartoon stimuli (Toonotopy) (Finzi et al., 2021) while performing a color-change detection task at fixation. Data were used to define each voxel’s spatial pRF parameters with the Viistasoft toolbox, visual areas in ventral, dorsal, and lateral pathways, and voxels with pRFs overlapping stimuli in the Sim-Seq experiment.

Session 2: Sim-Seq experiment. Observers viewed colorful cartoon segments in the upper left and lower right quadrants (centered at 5° eccentricity) while performing a 1-back RVSP task at fixation. In each quadrant, four squares were presented either sequentially (one at a time and in random order), or simultaneously (all at once). To examine how spatial and temporal summation contributes to fMRI responses, we collected data with high temporal resolution (TR=1s), where both simultaneous and sequential conditions used two stimulus durations (short: 200 ms and long: 1000 ms) and two stimulus sizes (small: 4 deg² and big: 16 deg²). Stimuli were shown in 8s blocks interspersed by blanks. Blocks were pseudo-randomized and repeated 16 times across 8 runs.

pRF modeling framework

Our pRF modeling framework has two stages. First, we build pRFs with spatial parameters from Toonotopy data and temporal parameters from prior work (Stigliani,
Jeska, & Grill-Spector, 2019). Once pRFs are defined, we select pRF centers overlapping Sim-Seq stimuli. Second, we predicted pRF responses for Sim-Seq stimuli at ms time resolution using three pRF models. Specifically, we tested if suppression may be due to (i) spatial linear summation (Dumoulin & Wandell, 2008), (ii) compressive spatial summation model (CSS; Kay, Winawer, Mezer, & Wandell, 2013), or (iii) compressive spatiotemporal summation model (CTS; I. Kim, Kupers, Lemme-Usabiaga, Shim, & Grill-Spector, 2022).

The newly developed CTS model binarizes the stimulus and applies 3 spatiotemporal filters: sustained, on- and off-transient. Filter outputs are rectified and go through a compressive static nonlinearity (power-law exponent). We use linear regression with split-half cross-validation to fit each voxel’s scale factors (β) and CST exponents using half the runs, and compute coefficient of determination (R²) for left-out runs.

**Results**

**V1 sums linear in space, but nonlinear in time**

V1 data show similar responses for simultaneous and sequential conditions, and larger for big vs small squares (Fig 1A). This response pattern is predicted by the linear pRF model, as pRFs are small and typically cover one square (Fig 1B, inset). Interestingly, many V1 voxels show a temporal nonlinearity: larger responses for short vs long durations (Fig 1A,B). Linear and CSS models cannot capture this effect—they predict the opposite: larger responses for longer durations. The CST model can predict this nonlinearity due to its transient spatiotemporal channels and hence captures most variance in the data (Fig 1C).

**Suppression, nonlinear spatial, and nonlinear temporal summation beyond V1**

In V2 and high-level areas, we find that single voxel responses are lower for simultaneous than sequential stimuli, higher for shorter than longer durations, and do not increase much with stimulus size (e.g., hV4; Fig 2A,B). Although pRFs in these regions are large and cover multiple squares (Fig 2B, inset), the linear pRF model failed to predict our observations. While both CSS and CTS pRF models predicted suppression and modest increase with stimulus size, the CSS model often overpredicted simultaneous suppression, that is, simultaneous responses were predicted to be lower than observed. Additionally, the CSS pRF model did not predict larger responses for shorter durations. The CTS pRF model could predict all three effects, including the increase in responses for brief stimuli (Fig 2B).

Comparing all three pRF models, CTS explained most variance in the fMRI data across the visual hierarchy (e.g., hV4; Fig 2C).

![Figure 2: hV4 results](image)

**Figure 2:** hV4 results (A) Average BOLD amplitude shows simultaneous suppression and stronger responses for shorter stimuli for one subject; (B) Example voxel pRF data (shaded) and CST model fit (dashed) where pRF is large and overlaps all four stimuli in quadrant; (C) Variance explained by each pRF model across 10 subjects.

**Discussion**

We demonstrate a new computational framework that examines the neural mechanisms of sensory suppression in human visual cortex. Our CST model provides an exciting, parsimonious explanation for the observed neural spatial and temporal nonlinearities via compressive spatiotemporal summation at the sub second range. This pRF modeling approach offers new insights into neural computations of dynamic visual inputs across visual hierarchies, as can be applied to other sensory (e.g., audition) and cognitive (e.g., attention) domains.
Acknowledgments
This research was supported by the US NIH NEI R01 EY023915 (KGS). Funders had no role in study design, data collection and analysis, or decision to submit the abstract.

References


